Understanding User Tradeoffs for Search in Encrypted Communication

Wei Bai, Ciara Lynton, Charalampos Papamanthou and Michelle L. Mazurek
University of Maryland, College Park
Email: {wbai, clynton, cpap, mmazurek}@umd.edu

***Abstract*—End-to-end message encryption is the only way to achieve absolute message privacy. However, searching over end-to-end encrypted messages is complicated. Several popular instant messaging tools (e.g., WhatsApp, iMessage) circumvent this inconvenience by storing the search index locally on the devices. Another approach, called *searchable encryption*, allows users to search encrypted messages without storing the search index locally. These approaches have inherent tradeoffs between usability and security properties, yet little is known about how general users value these tradeoffs, especially in the context of email rather than instant messaging. In this paper, we systematize these tradeoffs in order to identify key feature differences. We use these differences as the basis for a choice-based conjoint analysis experiment focused on email (n=160), in which participants make a series of choices between email services with competing features. The results allow us to quantify the relative importance of each feature. We find that users indicate high relative importance for increasing privacy and minimizing local storage requirements. While privacy is more important overall, local storage is more important than adding additional marginal privacy after an initial improvement. These results suggest that local indexing, which provides more privacy, may often be appropriate for encrypted email, but that searchable encryption, which limits local storage, may also hold promise for some users.**

1. Introduction

As people rely more and more heavily on online communication, privacy is becoming increasingly important. End-to-end encryption, in which the communications service cannot read the user's content, is the only way to fully protect users' online communications from malicious attackers, rogue company employees, and government surveillance. In recent years, more online communication services, especially instant messaging tools, have adopted end-to-end encryption; for example, WhatsApp and Apple's iMessage have brought end-to-end encrypted messaging to millions of users by default [1], [2]. While these popular tools have been moving in the right direction for securing communication, email adoption has not caught up [3].

While end-to-end encryption provides important privacy benefits, it can complicate functionality in a variety of ways that can potentially inconvenience end users. As one example, message searching (i.e., recovering a previously sent or received message using keywords) is a critical function for many users, especially for email [4]–[6]. When messages are end-to-end encrypted, searching their contents becomes complicated. Straightforward solutions in which the communication service maintains a central searchable index cannot be applied directly, as the communication service cannot read and index messages. To avoid this, instant messaging systems typically use a local index, in which messages decrypted on a device are indexed (and can therefore be searched) on that device.

An alternative to local indexing is to support searching over encrypted content. During the past two decades, researchers have made significant progress on *searchable encryption*, or mechanisms that can support searching encrypted data on an untrusted server without revealing the content of the messages or of the search queries [7]–[12]. These techniques are gradually being adopted by industry, e.g., for databases. While they have not yet been applied specifically for messaging, researchers have begun to explore this potential use case.

Both of these approaches to enable search for end-to-end encrypted messaging have benefits and drawbacks, in terms of security properties and in terms of utility for end users. To our knowledge, these tradeoffs have not been systematically explored from the point of view of end users. In this work, we take a first step toward filling that gap. First, we systematize the tradeoffs of these solutions in four dimensions: search expressiveness, performance, portability, and security across devices. This approach allows us to identify key features (that apply to both messaging and email) that may influence users' preferences. We then apply this systematization to design a choice-based conjoint analysis study focusing on email. We choose to study email specifically because of the prominence of search in an email setting, as well as the tendency for email archives to be relatively large and long-lived. Study participants were asked to make a series of choices between email-service *profiles* with different usability and security features; the results allow us to quantify the importance of each feature relative to the others in the email setting [13]–[15]. This methodology has been adopted in previous research to investigate users' online privacy concerns [16]–[18].

Our experiment (n=160) compared six features that may affect users' preferences for encrypted communication: price, multi-word search, partial-word search, privacy, local storage and synchronization across devices. We find that privacy is the second-most important feature, after price: participants indicated they would pay on average $0.85 per

month to reduce service providers' access to the contents of their messages and search queries. However, minimizing the use of local storage was almost as highly valued. We also find that participants' web skills, along with their self-reported level of privacy concern, influence their choices.

Our results support the use of local indexing in many cases to maximize privacy. However, our results also suggest that searchable encryption for encrypted email may be a promising avenue to support those users who want more privacy than in an unencrypted service but who also highly prioritize limiting use of storage on mobile devices.

We make the following concrete contributions:

- We systematize privacy and usability tradeoffs inherent in different approaches to adding search to end-to-end encrypted communication.
- We conduct a conjoint analysis study to better understand how end users value these various tradeoffs for email.
- Based on these results, we make recommendations for the design of end-to-end encrypted email systems and for further research.

2. Related work

In this section, we discuss related work in four key areas: adoption and usability of end-to-end encryption; advances in searchable encryption; habits of email search and organization; and applications of conjoint analysis.

2.1. End-to-end encryption, adoption, and usability

End-to-end-encrypted email communication was made more accessible by the release of PGP [19], which led to the development of several encrypted email systems (e.g., [20], [21]). Since 2004, the Off-The-Record protocol has been used as the basis for several tools for secure instant messaging, including Signal [22] and WhatsApp [1]. Unger et al. surveyed secure communication tools in 2015, and evaluated their security, usability and ease-of adoption properties [23].

During the past two decades, researchers and practitioners have made a considerable effort to understand problems with deployment of end-to-end encrypted communication tools and make them more accessible to users. Since the seminal work by Whitten and Tygar that identified the user interface design flaws in PGP 5.0 [24], researcher have explored several approaches for improving user interface designs, such as mitigating problems identified in previous systems [25], [26], using understandable metaphors [27], and designing more informative indicators [28]. Other studies also explored the influence of automatic message encryption and decryption [29] and integration with existing systems [30].

Many researchers have focused on the problems of key exchange and management in end-to-end encrypted systems [31]–[36]. Garfinkel et al. evaluated *key continuity management* (KCM), and concluded that it was a workable model. Fahl et al. proposed Confidentiality as a Service (CaaS), which split the trust between the communication service provider and the CaaS server, and found that user study participants could successfully use this approach to encrypt Facebook conversations [32], [35]. In pursuit of less cumbersome and more secure approaches to key management, Ryan extended the concept of *certificate transparancy* to end-to-end encrypted emails, and CONIKS allows key owners to monitor how their keys are distributed by a central key server [33], [34]. Bai et al. investigated general users' attitudes toward tradeoffs between security and usability of different key management models, finding that models which are more convenient but less secure may be perceived as "good enough" for everyday usage [36]. Lerner et al. proposed *Confidente*, a prototype that used *Keybase* [1] to centrally manage public keys and also possibly private keys protected by users' passwords [37]. They claimed that lawyer and journalist participants made fewer mistakes using *Confidente* made fewer mistakes than Mailvelope. Some other studies evaluated modern end-to-end encrypted systems and found that users were still susceptible to Man-in-the-Middle (MitM) attacks due to human errors [38], [39].

Social factors and network effects also play an important role in adoption of secure messaging. In 2007, Solove found users were reluctant to encrypt their emails because they "have nothing to hide" [40]. Similarly, Gaw et al. in 2006 found that even employees in a sensitive advocacy organization regarded routine email encryption as "paranoid" and socially undesirable [41]. More recently, researchers found that security and privacy played a minor role for general users to adopt secure messaging tools [42], and journalists' adoption of secure tools was driven by journalistic sources and existed tools didn't comply with their requirements [43]. Abu-Salma et al. found that for general users, usability was not a primary obstacle to adoption; instead, fragmented user bases, lack of interoperability, overall feature sets, and limited understanding of security properties are significant barriers [3]. These studies suggest that to promote adoption of encrypted messaging, the community must consider not just the usability of security features themselves, but rather the overall ecosystem of features and tradeoffs within which users make adoption decisions. This study contributes to this goal by examining how best to incorporate the critical feature of search into end-to-end-encrypted email systems.

2.2. Searchable encryption

Searchable encryption was first introduced in 2000 by Song. et al., using a symmetric encryption scheme [7]. Since then, numerous works have improved search efficiency and protocol security [8]–[12], [44], [45]. Many companies have integrated searchable encryption capabilities into products, primarily database systems [46]–[50]. Fuller et al. provide an overview of searchable encryption solutions from academia and industry, characterizing the tradeoffs between different solutions [51].

1. https://keybase.io

Although many searchable encryption solutions have been proposed and evaluated, very few studies have been done to understand users' perceptions toward such solutions. To our knowledge, the only such study is a 10-participant pilot that evaluated users' experiences using a web interface to query searchably-encrypted databases [51]. We complement this preliminary work by examining how users balance utility and privacy tradeoffs, in the context of end-to-end-encrypted email systems.

2.3. Email search, organization, and mobile usage

Searching emails, unlike other information retrieval tasks such as web search, is to refind "stuff I have seen" [52]. Researchers have spent significant effort to understand how users search their emails and help them to do so more effectively.

In a lab study, Elsweiler et al. explored some email search query characteristics on Mozilla Thunderbird, e.g. the query length and the field submitted on. In a followup field study, Harvey et al. discovered some search behavioral patterns in email, such as whether and how often the same or similar queries were submitted by the same user, and when users clicked through emails after users performed a search [53]. Whittaker et al. found that preparatory behaviors such as creating folders are inefficient and do not improve email retrieval success [54]. Cecchinato et al. found that users searched their emails in personal and work account differently due to different email management strategies [4].

More recently, researchers have been able to examine search queries from large email providers and identify broad patterns. Carmel et al. found that among Yahoo! Mail users, most search queries are fully formulated by users (rather than suggested by the email service), and most contain only one word [6]. Similar query lengths were found in Outlook queries [5]. Among the outlook queries, 18% contained "advanced" search operators, primarily the "from:" and "to:" operators used to specify email senders and recipients.

Users are increasingly accessing emails via mobile devices rather than desktops or laptops, reaching 53% in 2015 [55]. Email marketing reports also indicate that up to a quarter of users always read email on mobile first [56].

These email habits will affect users' preferences for searching encrypted as well as plaintext email, and they informed the design of our choice-based analysis.

2.4. Conjoint analysis for valuating privacy

Conjoint analysis is a statistical method for understanding users' relative preferences among multiple-attribute products or services [13]–[15], and researchers have adopted it to understand various aspects of users' online privacy preferences. Krasnova et al. used computer-aided Adaptive Conjoint Analysis (ACA) to investigate users' privacy values in online social networks (OSN) [17]. They expressed users' preferences for different attributes, including privacy, in monetary value, which allowed them to compare preferences among different features easily. In our study, we also adopt monetary values as a basis of comparison. They found that privacy was indeed important to OSN users. They also classified participants into three groups, and each group had different tradeoffs between privacy and other OSN features. Burda et al. explored users' cloud-storage choice decisions by employing a choice-based conjoint analysis [18]. They found that users preferred client-side encryption over server-side encryption, and estimated providing client-side encryption is equivalent to between 0.86 and 1.66 Euros per month. Pu et al. conducted both a full-profile and a choice-based conjoint analysis to understand the factors that influence social app users' valuation of their own and their friends privacy, measured by how much data was collected by an app [16], [57]. They found that app users valued all of their friends' privacy smaller than their own's [16], but vice versa in [57]. In this work, we apply choice-based conjoint analysis to examine users' value tradeoffs among utility and privacy in searching encrypted email.

3. Design space for search in encrypted communication

In this section we characterize security and usability tradeoffs for searching encrypted communications.

We first consider a **cloud index** (CI), defined as follows. When a new message arrives at the user's device, it is decrypted locally and then tokenized. Using a secret *tokenization key* stored on the local device, these tokens are encrypted, and a mapping between tokens and the associated message identifier is uploaded to the cloud-based search index. To search for messages containing a keyword w, the user's device uses the tokenization key to generate an encrypted token for w and sends it to the server. The server looks this token up in an index and returns all associated message identifiers. This approach is based on searchable encryption techniques such as those described by Stefanov et al. and Bost [11], [12].

We contrast this approach with a **local index** (LI). When a new message arrives, it is decrypted locally, and its contents are added to an unencrypted search index stored on the local device. All queries are handled locally, so no information about the index or queries is provided to the communications service. This is the approach used by WhatsApp and iMessage.

As a basis of comparison, we also include two control approaches: **no end-to-end encryption** (None) (as in most popular email services today) and **end-to-end encryption without search** (NoSearch) (as in services like Mailvelope [21]).

We do not include ORAM-based schemes, as they are not yet sufficiently practical in terms of performance.

We compare our four chosen approaches on the following metrics:

- Expressiveness: The types of search expressions supported by each solution.

	None	NoSearch	CI	LI
Expressiveness				
Exact single word	✓	—	✓	✓
Multi-word boolean	✓	—	✓	✓
Exact string	✓	—	—	✓
Partial word	✓	—	—	✓
Aux. expressions:				
Sender/recipient	✓	✓	✓	✓
Subject line	✓	—	✓	✓
Date	✓	✓	✓	✓
Label or folder	✓	—	✓	✓
Performance				
Server storage	$\mathcal{O}(N)$	—	$\mathcal{O}(N)$	0
Client storage	0	—	$\mathcal{O}(WlogD)$	$\mathcal{O}(N)$
Bandwidth	$\mathcal{Q}+\mathcal{L}$	—	$\mathcal{Q}+\mathcal{L}+\mathcal{U}$	0
Latency	low	—	low	low

TABLE 1: Supported search expressions and performance for different search approaches. For expressions, ✓ indicates that an expression type is supported; expression types that cannot be effectively supported are indicated with "—". We note that current industry products do not necessarily support all types of expressions; this analysis is based on capabilities rather than implemented products. For performance, $\mathcal{Q}$, $\mathcal{L}$ and $\mathcal{U}$ mean the size of query content, returned list of identified items, and update information, respectively.

- Performance: The cost of searching and of updating the search index in server storage, client storage, bandwidth consumption, and time.
- Portability: How well each solution can be scaled to multiple devices, including initializing a new device and routinely switching devices.
- Security: Threat models each solution can and cannot defend against.

3.1. Expressiveness

In this section, we consider the types of search expressions each approach can support. We surveyed current mainstream email and instant messaging systems, including Gmail, Outlook, Apple Mail, Yahoo Mail, WhatsApp and iMessage, to create the following list of currently supported and widely used search techniques:

- Exact single word. Users search for one word, and get back messages containing exactly that word.
- Multiple-word boolean. The search expression contains more than one keyword. By default these are typically combined via AND, but OR and NOT are also potential options.
- Exact string. A query that returns all messages that contain the exact multi-word string, in order. This is often implemented as a query string enclosed in quotation marks.
- Partial word. The system returns any messages containing words that are superstrings of the provided keyword.

	None	NoSearch	CI		LI	
			Basic	Password	Basic	Password
Portability						
Old_msg	Yes	No	No	Yes	No	Yes
Index	Shared	—	Own	Shared	Own	Inherited
Security						
Keyword	All	None	More	$\mathcal{P}$	Less	$\mathcal{P}$
Message	All	None	More	$\mathcal{P}$	Less	$\mathcal{P}$

TABLE 2: Portability and security for variations of the CI and LI approaches. For portability, the first row concerns whether old messages (from before device activation) can be read and searched on new devices. The second row concerns how indexes are constructed — independently per device ("own"), "shared" across devices, or "inherited" from another device. For security, the extent to which search keywords and messages is indicated (on a relative scale, in order) as "All", "More", "Less", and "None." $\mathcal{P}$ means security depends on the password strength.

- Auxiliary expressions. These expressions, which are sometimes referred to as advanced search operators, allow the user to specify that specific keyword should be found in a specific field, such as the sender or subject line; to specify a date range within which to search; or to specify a label or folder within which to search.

Our results are summarized in Table 1. In the None approach, the service has access to all plaintext messages and therefore can support all types of search expressions. In contrast, in the NoSearch approach the server cannot search over encrypted messages at all, so only searches based on the sender/recipient and the date are available. The date can always be observed by the communications service, and the sender and recipient cannot be easily encrypted if the message is to be correctly delivered.[2] If subjects and labels/folders are also encrypted, they cannot be searched over, although we observe that current tools, such as Mailvelope, do not encrypt subjects and labels/folders and therefore can be searched.

In the LI case, all information is obtained from plaintext emails and stored on the local device; as such, all the expressions we consider are supported. The nature of the CI approach inherently allows single keywords and their boolean combinations; however, boolean combinations leak more information about the user's search patterns and therefore their message contents [58]. As the number of keywords involved in these combinations increases, search performance in the CI approach degrades. Recent research measuring email usage patterns, however, suggests that the average number of keywords per email search query is only about 1.5, suggesting that the common case for multi-word boolean searching is reasonable for CI systems. Searchable encryption could support exact multi-word strings by indexing n-grams of various lengths; however, as this would de-

2. There are some email services that provide identity anonymity, e.g. TorGuard (https://torguard.net/anonymous-email.php), but this consideration is out of scope for this paper.

grade performance even more quickly, and strings of length greater than two or three words would quickly become impractical, we consider it unsupported.

We next consider whether CI can support arbitrary partial keywords. There are some partial solutions, but those proposed in [59]–[61] did not support updates, and both [62] and [63] adopted ORAM schemes. Some partial keyword search is possible if word stemming (e.g., Porter Stemmer [3]) is applied before message indexing, and the keyword is exactly the anticipated stem. Overall, we consider this unsupported.

As to auxiliary expressions, CI can support sender/recipient and date expressions for the same reasons that NoSearch can, described above. If subject lines and labels/folders are encrypted, they can be tokenized and searched just like message contents, so we say that these expressions are feasible. We distinguish this from the None case, where these items can be either encrypted or searched, but not both.[4]

3.2. Performance

We next consider performance metrics with which to evaluate each search solution. In particular, we consider the cloud and local storage space required, bandwidth consumption in search and update operations, and latency in search and update operations. The results are summarized in Table 1.

The notation used in this section is as follows: N denotes the total number of keyword/message pairs, W is the number of unique keywords, and D is the total number of encrypted messages. We use the performance reported in [12] as a reference example for a searchable encryption scheme, because it provides a good balance between performance and other metrics.

First we consider server and local storage requirements. For NoSearch, there is no index, and for None, the index is stored entirely in the cloud and is approximately proportional to N, adjusted for some data compression. For LI, the index is stored entirely locally, and is again roughly proportional to N, adjusted for data compression. For CI, the situation is more complex. Some local storage is required, roughly proportional to $W \log D$, but the bulk of storage is in the cloud. This storage is proportional to N, but larger than the storage required for None because of encryption-scheme overhead. In the Bost scheme, for example, each keyword/message pair requires 32B of cloud storage [12].

We next consider bandwidth consumed during search queries and updates. Trivially, NoSearch consumes no bandwidth; because LI is purely local it similarly consumes none. None consumes bandwidth equal to submitting the query content and returning the list of identified items. CI operates similarly, but with additional overhead caused by the encryption scheme. CI also requires the client to send update information in order to index a new message. Bost's scheme requires 16B per input token and 32B per output keyword/message pair during search, along with 32B per newly indexed keyword/message pair.

Finally, we consider latency. CI is more time consuming, both in search and in updates, than the LI and None approaches, because it requires additional computation. However, overall the latency performance of such schemes has become very fast. Microbenchmarks reported in [12] demonstrate that searching among 140 million keyword/message pairs, with 10 results returned, requires only 24 μs. This latency will increase as the number of keyword/message pairs — and in particular, the number of unique keywords — increases. Updates, upon receiving and indexing a new encrypted message, can generally occur in the background rather than while the end user is waiting; Bost reports an update throughput of around 4,300 keyword/message pairs per second.

3.3. Portability

Previous studies have shown that the number of devices people use to access their email has grown over the past few years [4], [64]. Therefore, it is important to consider how each solution can be scaled to more than one device.

None solutions, of course, are trivially portable. The portability of NoSearch, CI and LI solutions depends primarily on the underlying key management model, as shown in Table 2. In the basic model, as adopted by Mailvelope [21], the key pair is generated locally and the key is not shared with other devices. As such, when a new device is added to the ensemble, it does not have access to messages that were not encrypted for the new device's key. For both CI and LI, this means that devices cannot read or search messages from before device activation. Each device requires its own separate encrypted index, and upon activation the new device begins building its index from scratch.

An alternative model applies a password-protected approach to share keys. We refer to this as the *password-protected* variation. This model allows users to decrypt old messages on new devices by retrieving corresponding keys from the cloud, and it has been adopted by several existing products or research prototypes, such as *Confidante* [37], ProtonMail [65], and Threema [66]. For CI, sharing a password-protected key allows new devices to read and search all old messages, and allows different devices to share one index in the cloud. In the password-protected LI, a new device can access old messages and can inherit an existing index and update it as new messages are received.

There may be other portability solutions as well, but we consider these basic and password-protected approaches as exemplars for our analysis.

3.4. Security

This section briefly describes the threat model each solution can defend against, as well as its weaknesses. We focus

3. https://tartarus.org/martin/PorterStemmer/

4. We note that email systems may not be able to create and populate labels or folders automatically because they cannot read email contents.

on the ability of service providers (e.g., email companies) to learn their users' message contents and search queries. We do not address interception by third parties or endpoint compromise, as these are for the most part orthogonal to the implementation of search. We assume a semi-honest scenario, where the provider is curious but does not deviate from protocols.

We first consider our two controls, None and NoSearch. In None solutions, since all messages are in plaintext, the provider will learn all message contents as well as all of a user's search queries. This is how most email services currently work (e.g., Gmail [67]). We regard this solution as the lower bound for security. In NoSearch solutions, only the user can read messages sent to her, and all other parties, including the provider, cannot. There are no search queries, so the provider cannot learn them. We consider this solution the upper bound for security in our evaluation.

Table 2 shows some security features for CI and LI. Many different CI solutions provide more security than None, but less than NoSearch. Generally, the content of search queries is concealed but the search pattern, access pattern, and size pattern are revealed to the provider (for detailed definitions, we refer readers to [8]). Further, many proposed CI approaches achieve forward privacy, meaning that when a user searches for a keyword, if a later message containing the same keyword is added, the server cannot connect it to the prior search.

In the past few years, researchers have investigated how to exploit the information leaked by these schemes. CI solutions built from Bloom filters (e.g., [68], [69]) are potentially vulnerable to message recovery via leaked information [70]. There are also many proposed attacks for recovering search keywords, the details of which differ based on assumptions about the providers' abilities as well as about the composition and distribution of messages and search keywords. For example, providers who can send encrypted messages to the client, potentially from a spoofed address, can learn whether a user is searching within a set of candidate keywords [71].

For LI solutions, leaked information is minimal, as all searches are performed locally. However, eliminating leakage entirely has been proven impossible [72]; for example, the provider may observe patterns in which old messages are requested from the server (presumably after a local search has completed). Overall, we consider LI to be stronger than CI but not precisely as secure as NoSearch.

If we consider the special case of password-protected key and/or index material discussed in Section 3.3 above, the security of both CI and LI reduces to the strength of the user's password.

4. Email preference study

In the previous section, we identified several key tradeoffs related to search in encrypted communication systems. Choosing among these tradeoffs, however, requires understanding how users balance and prioritize different features. As a first step toward understanding this, we conducted a *choice-based conjoint analysis* study focused on email. In conjoint analysis, participants are presented with a set of product *profiles* that vary in several dimensions. By choosing among them, participants reveal their relative preferences among the various features [73], [74]. In this study, we asked participants to choose among email services with different features drawn from the tradeoff analysis presented in Section 3. In this section, we describe the features we selected, then detail the design of our study and our analysis approach.

4.1. Features and options

For the rest of the paper, we define *feature* as a general property of a choice profile — for example, price or security posture — and we define *option* as one of various settings a given feature can take. For example, the price feature has options of $0.00 and $1.99.

Based on the tradeoffs identified in Section 3, we identified six features to examine: price, multi-word email search, partial-word email search, local (device) storage, portability of old messages to a new device, and privacy. Each feature has two or three possible options. All features and options are listed in Table 3.

The first feature we include is price; this is included because it is familiar to participants when thinking about value tradeoffs, and because it allows us to express the resulting valuations for different features in dollars. We chose to use a monthly fee to make the price a bit more meaningful for users on an ongoing basis. To set the price, we searched "secure email" in both Apple Store and Google Play, and looked at the top 50 returned apps; more than 90% were priced at $1.99 or lower, and the majority are free. Therefore, we set two options for the price feature: $0.00 and $1.99. We note that setting the best possible price is not critical, as our main goal is to choose something differentiable enough to matter to users, but not so expensive that it overwhelms other features.

We chose two expressiveness features: multi-word search and partial-word search. These features are used relatively often for email search [6], [53], and they vary between the two main search solutions we consider. (For simplicity, we presented users with only "multi-word search" and did not differentiate between boolean and exact string matching.) We do not include single keyword, or auxiliary expressions that consider the sender/receiver or date, as they are supported by both LI and CI. More complex auxiliary expressions are also rarely used [5]. In the conjoint analysis, each expressiveness feature has two options: "yes" (available) or "no" (unavailable).

As discussed in Section 3.2, the most notable performance difference among approaches we consider is related to storage on the local device. The storage required, of course, depends on how many keyword-email pairs a particular user generates, and will increase over time as more messages are added. For simplicity, however, we set two possible options for the storage feature: 5MB of local storage (low storage) and 500MB of local storage (high storage). The

Features	Feature Descriptions	Options
Price	Monthly fee for signing up for the service	**$0.00 (free)** **$1.99 per month**
Multi-word Email Search	When you search emails, can you search using multiple words, for example, "home depot"?	**Yes** **No**
Partial-word Email Search	When you search emails, can you search with a partial word instead of a complete word, e.g., "amaz" vs "amazon"	**Yes** **No**
Storage on your device	How much local storage the service will use (on your computer or phone)	**Will use 5MB on your device**, equivalent to about 2-3 HD photos **Will use 500MB on your device**, equivalent to about 200 HD photos or 20 mobile apps
Syncing old email to a new device	After using the email service on one device (a phone or laptop) for several months, you buy a new device. You set up the email service on your new device. Can you read and search old emails on your new device?	**Yes – read and search all email** using the new device both before and after its activation. **No – read and search only email** after you configured the new device. You can't read and search old emails on your new device.
Privacy	What can your email service learn about your email and email search queries?	**Standard privacy**: Your email service can access the contents of all email and email search queries. This is how most email services currently work (Gmail, Yahoo! mail, Outlook, etc.) **Extra privacy**: Your email service can access the contents of certain emails and email search queries, but not all. The service can choose specific topics of high interest to learn about. **Maximum privacy**: Your email service cannot access the contents of any emails or email search queries.

TABLE 3: Features and options used in our conjoint analysis experiment. We selected six features, with two to three possible options for each.

5MB figure represents CI and is adapted from Bost, reflecting 14 million keyword-message pairs and 213,349 unique keywords [12]. The 500MB figure is intended to represent LI and was selected as a round number roughly in line with the cloud storage requirements for the same Bost scenario (because for LI, all index data is stored locally instead of in the cloud). To make this size difference more intuitive for participants, we also describe the storage requirements in terms of the equivalent number of HD photos or apps.

For the portability feature, we selected the ability to work with old emails on new devices. This feature distinguishes the basic CI and LI approaches from their password-protected variations, and is relatively easy to explain. We describe this feature to participants as follows: "After using the email service on one device (a phone or laptop) for several months, you buy a new device. You set up the email service on your new device. Can you read and search old emails on your new device?" The possible options are "yes" and "no."

The final feature we consider is privacy. For ease of explanation to users, we describe the privacy feature as the ability of the email provider to access users' email contents and email search queries. Although the actual security properties of various searchable encryption schemes have subtle differences, for our users we simplify this feature into three options: *standard* privacy, *extra* privacy, and *maximum* privacy. Standard privacy describes a scenario with no end-to-end encryption, as is common in most email services today. (We found in pilot testing that summarizing this option as "low" or "no" privacy unduly alarmed users, so we chose standard to reflect the current common case.) The extra privacy option was described as allowing the provider to access "some but not all" message contents and search queries, with the provider having some choice of topics to learn about. This category was designed to approximate CI solutions. Lastly, maximum privacy was described as the provider being unable to learn any email or search querty contents. This option was designed to represent NoSearch as well as LI.

By asking users to choose among various combinations of these feature options, we can determine the relative value of each feature compared to the rest.

4.2. Study setup: Choice-based analysis

The feature set described in the previous section totals six features, with up to 96 ($2 \times 2 \times 2 \times 3 \times 2 \times 2$) possible option combinations for a single profile. In a traditional full-factorial, full-profile conjoint analysis, participants would be asked to rank all 96 of these possible profiles; clearly this is an untenable cognitive burden [74]. To address this, we elect to use a choice-based analysis, in which participants make a series of discrete choices between two profiles, rather than ranking a large set of profiles [75]. We also reduce the set of profiles we examine by using a fractional-factorial design that consists of a smaller number of orthogonal profiles [76]. Prior work has found that participants can handle as many as 17 choice sets without problems [77]. We therefore construct a set of 16 distinct choice pairs. To choose the 16 distinct choice pairs, we generally follow the randomized approach suggested in [78]. A screenshot of a single choice pair is shown in Figure 1, and the screenshots of all 16 questions used in our study are shown in the appendix.

It is common in conjoint analysis to provide a "no choice" option, indicating that the participant will choose neither of the presented profiles. In this case, however, we are primarily interested in learning the relative value of the different features, so "no choice" options (indicating that a participant would prefer to keep her existing email service) are somewhat unhelpful. We instead provide a "no

preference" option to allow a participant to indicate that the two choices are (to her) equivalent.

4.3. Protocol and recruitment

We recruited participants from Amazon's Mechanical Turk crowdsourcing service [79] who are 18 or older and fluent in English. To improve data quality, we required participants to have completed at least 50 prior Human Intelligence Tasks (HITs) with a HIT approval rating over 95% [80]. Because perceptions of dollar value and privacy preferences are likely to vary across countries, we restrict recruitment to Turkers in the United States. We advertised our task as asking participants' preferences about email services, and explicitly did not mention privacy to avoid self-selection bias. We prevented duplicate participation based on MTurk ID. Participants were paid $1.50 for the study, which was expected to take less than 15 minutes. On average, it took participants around 12 minutes to finish our questionnaire.

Study participants first received general instructions explaining that they would be required to answer questions about recommending a new email service to a friend. These instructions included the overall feature table (Table 3). To encourage participants to read these instructions carefully, we enforced a six-second delay before the participant could advance to the next screen. We also required participants to check a box agreeing that they would "read the instructions and each question carefully and think about your answer before you give it. We rely on our respondents being thoughtful and taking this task seriously."

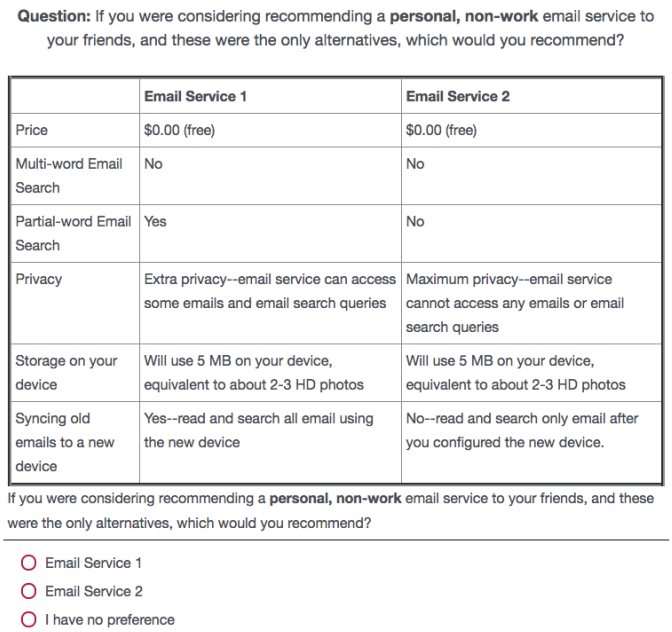
The participant was then shown the choice-pair questions, one at a time, as shown in Figure 1. We presented the overall feature table above each question for the participant's ongoing reference.

After the choice-pair questions, we asked participants to answer knowledge, attitude, and demographic questions. These included the six-item abbreviated web-use skills index [81], which measures web-based tech savviness; the 10-item Internet Users' Information Privacy Concerns (IUIPC) [82], which measures consumer privacy concern; and standard questions about gender, age, ethnicity, education, and income. Several privacy concern scales have been developed, each of which has various benefits and drawbacks; the IUIPC is recommended by Preibusch in his survey [83].

Our study protocol was approved by the University of Maryland Institutional Review Board (IRB).

4.4. Data analysis

We used Hierarchical Bayesian (HB) models to analyze choice-based conjoint data. Compared with other statistical analysis methods, HB models are able to estimate participants' preferences not only at the aggregate level (i.e. general preferences from participants), but also at the individual level [84]. We applied the R package "bayesm" [85] to conduct this HB analysis.



Question: If you were considering recommending a **personal, non-work** email service to your friends, and these were the only alternatives, which would you recommend?

	Email Service 1	Email Service 2
Price	$0.00 (free)	$0.00 (free)
Multi-word Email Search	No	No
Partial-word Email Search	Yes	No
Privacy	Extra privacy--email service can access some emails and email search queries	Maximum privacy--email service cannot access any emails or email search queries
Storage on your device	Will use 5 MB on your device, equivalent to about 2-3 HD photos	Will use 5 MB on your device, equivalent to about 2-3 HD photos
Syncing old emails to a new device	Yes--read and search all email using the new device	No--read and search only email after you configured the new device.

If you were considering recommending a **personal, non-work** email service to your friends, and these were the only alternatives, which would you recommend?

- ○ Email Service 1
- ○ Email Service 2
- ○ I have no preference

Figure 1: Screenshot of a choice question example

In our HB analysis, we used *effects coding* method for categorical coding of feature options [86]. In effects coding, one option of each feature is chosen as the *baseline*, to which other options are compared. In our analysis, we selected the least desirable option for each feature ($1.99, no multi-word search, no partial-word search, standard privacy, 500MB local storage occupied, and no sync ability) as baselines.

After our conjoint analysis, we applied linear regression to estimate the effects of participants' web-use skill scores and three IUIPC scores (collection, awareness, and control) on their marginal valuation of privacy at different levels. In order to prevent possibly over-fitting models, we applied the standard backward-elimination model selection process, until the Akaike Information Criterion (AIC) is minimized [87]. For each regression model, we present what variables are selected, estimated coefficients and p-values.

4.5. Limitations

Our study has several limitations. Amazon MTurkers are younger, better educated and more technical than the general U.S. population [88], which may limit the generalizability of our results. Our participants, though, (as detailed in Section 5.1) showed similar online privacy concerns to the IUIPC scores reported in [82].

Participants' self-reported preferences don't always match their revealed preferences [89], [90]. This means that overall, we expect privacy to be overvalued a bit in our results. However, requiring participants to make explicit tradeoffs (rather than simply asking about whether privacy is important) may mitigate this somewhat. We asked participants to recommend email services to their friends instead of using themselves, to make questions more neutral and less

personally fraught. Although participants' recommendations may not always align with choices of their own use, this is a known technique for surveying about sensitive topics [91] and has been used when investigating app permissions [92]. Overall, we believe our results can provide a meaningful jumping-off point for understanding how users value trade-offs between security and usability features in searchable encrypted messaging tools, but further work is needed to examine revealed preferences.

To avoid overwhelming our participants, or requiring them to understand complex technical distinctions, we presented participants with only three coarse levels of available security. In reality, various searchable encryption solutions have subtly but meaningfully different security guarantees [51], and these subtleties are not captured in our results. Future work can examine these alternatives in more detail.

More generally, we asked participants to make nuanced decisions that required reading and thinking about multiple features and option levels. Some participants may have answered carelessly, or used satisficing rather than thinking deeply about each option [93]. To mitigate this, we restricted MTurk recruitment (as described above), we kept the study short, and we iteratively tested our feature descriptions to make them as clear as possible.

5. Results

In this section, we present the demographics of our participants, their preferences among features and options, and the correlation between their privacy preferences and their online privacy concerns and web skills.

5.1. Participants

In May 2017, we collected 160 completed questionnaires. Self-reported demographics for the those 160 participants are shown in Table 4, which shows that 51.3% of our participants were male, 41.0% were in the age range of 30-39, 80.1% were Caucasian, and 43.6% held a bachelor's degree. In addition, 78% participants held neither a degree nor a job in an IT-related field, and 30% reported household income in the $50k-$75k range in 2016.

We also measured participants' web knowledge and online privacy concerns using two extensively validated scales. To assess web knowledge and skills, we adopt the six-item web-use skills index for the general population, developed by Hargittai et al. [81], which measures web skills on a Likert scale from 1 (low) to 5 (high). Our participants reported an average score of 4.2 (SD = 0.80). This is higher than the 3.4 average reported by Hargittai et al., which we attribute to MTurkers having higher than average internet skills and to population changes since the Hargittai data was collected in 2010.

To measure internet privacy concern, we adopt the 10-item IUIPC scale developed by Malhotra et al. [82]. There are several available scales for measuring privacy, each of which has known weaknesses; the IUIPC scale that we use is recommended by Preibusch as "a safe bet," but should be interpreted carefully [83]. On a seven-point scale, with seven indicating highest privacy concern, our participants averaged 5.9 (SD=0.81) for consumer control, 6.3 (SD=0.66) for awareness of privacy practices, and 5.61 (SD=1.07) for data collection. These results are comparable to those Malhotra et al. reported: 5.7, 6.2, and 5.6 respectively.

Gender	Male	51.3%
	Female	48.1%
Age	18-29	21.2%
	30-39	41.0%
	40-49	19.9%
	50-59	8.3%
	60-69	7.7%
	70+	1.9%
Ethnicity	Caucasian	80.1%
	Hispanic	3.8%
	Asian	9.0%
	African American	6.4%
Education	Completed H.S. or below	8.4%
	Some college	22.4%
	Associate's degree	11.5%
	Bachelor's degree	43.6%
	Master's degree or higher	9.7%
IT-related Job or degree	Yes	18.8%
	No	78.2%
Income	<$30k	23.7%
	$30k-$50k	25.6%
	$50k-$75k	30.0%
	$75k-$100k	9.4%
	$100k-$150k	7.5%
	$150k+	1.9%

TABLE 4: Participant demographics. Percentages may not add to 100% due to "other" categories and item non-response.

5.2. Results for conjoint analysis

We present participants' preferences for different features and their options in this section. We follow the analysis procedures of Burda et al. [18]. We first calculated the individual and aggregated part-worth utilities for each option. Based on those utilities, we then estimate the relative importance of each feature, the utility changes between feature options and the corresponding monetary values in terms of dollars.

5.2.1. Model fitting. First, we evaluate how well our estimated HB model is fit. In order to evaluate this goodness-of-fit, we conducted the test suggested in [94]. We calculated the *likelihood ratio* (LR) to measure how well our estimated model performs compared with a dummy model in which all parameters are zero [94]. This test shows that our estimated model is statistically valid with LR = 25.89 (p=0.001). In other words, the hypothesis that our estimated model and the null model are equal can be rejected. Second, we calculated

Features	Options	Part-worth Utility	Part-worth Utility CI	Relative Importance
Price	\$0.00 (free) \$1.99 per month	1.856 -1.856	[1.851, 1.860] [-1.860, -1.851]	32.59%
Privacy	Maximum privacy Extra privacy Standard privacy	1.219 0.190 -1.409	[1.215, 1.223] [0.188, 0.192] [-1.413, -1.405]	24.19%
Local Storage	5 MB 500 MB	0.635 -0.635	[0.630, 0.639] [-0.639, -0.630]	17.38%
Sync old email to a new device	Yes No	0.410 -0.410	[0.409, 0.411] [-0.411, -0.409]	9.36%
Multi-word Search	Yes No	0.460 -0.460	[0.458, 0.463] [-0.463, -0.458]	9.02%
Partial-word Search	Yes No	0.392 -0.392	[0.391, 0.394] [-0.394, -0.391]	7.46%

TABLE 5: Part-worth utilities of feature options and feature relative importance. CI indicates 95% confidence interval.

Features	Option Change	Utility Change	Utility Change CI	Dollar Value	Dollar Value CI
Price	1.99 $\rightarrow$ 0.00	3.711			
Privacy	Standard $\rightarrow$ Extra	1.599	[1.438, 1.760]	0.86	[-0.05, 1.77]
	Extra $\rightarrow$ Maximum	1.029	[0.874, 1.185]	0.55	[-0.05, 1.16]
	Standard $\rightarrow$ Maximum	2.628	[2.351, 2.905]	1.41	[-0.01, 2.83]
Local Storage	500 MB $\rightarrow$ 5 MB	1.269	[0.928, 1.611]	0.68	[-1.12, 2.48]
Sync old email to a new device	No $\rightarrow$ Yes	0.819	[0.669, 0.970]	0.44	[-1.40, 2.29]
Multi-word Search	No $\rightarrow$ Yes	0.921	[0.781, 1.060]	0.49	[-0.21, 1.20]
Partial-word Search	No $\rightarrow$ Yes	0.785	[0.682, 0.888]	0.42	[-0.03, 0.87]

TABLE 6: Utility change in feature options and monetary values. CI indicates 95% confidence interval.

	\$0	Multi	Partial	Extra	Max	5MB	Sync
\$0	1.00	-0.13	-0.04	0.03	0.18	-0.13	-0.07
Multi		1.00	0.05	0.03	-0.01	-0.10	0.06
Partial			1.00	0.35	-0.09	-0.07	0.12
Extra				1.00	0.03	-0.02	0.07
Max					1.00	-0.13	0.06
5MB						1.00	0.04
Sync							1.00

TABLE 7: Correlation matrix of non-omitted options in HB analysis. Notes: "multi" means multi-word search supportive; "partial" means partial-word search supportive; "extra" and "max" mean extra and maximum privacy, respectively; "5MB" means 5MB local storage needed; "sync" means synchronizing supportive.

the hit rate in our 16 choice questions by identifying for each participant the *profile* with the highest probability based on the estimated model, and then determining whether or not that participant actually chose this *profile*. The test results in a hit rate of 89.84%, compared to 33% random guessing, suggesting that our model is well-fitted.

5.2.2. Interpreting the model. The outcomes of Hierarchical Bayesian (HB) models are called *part-worth utilities*. Part-worth utilities are a unitless measure of the relative value of different options within each feature. We use the *effects coding* method for quantifying feature options [86]; as such, for each feature, the part-worth utilities will sum to 0, with the least-desirable option showing a negative part-worth utility. Part-worth utilities are calculated independently for each participant, then averaged to produce an overall result.

More formally, the probability of participant n choosing email service profile j in question k is shown in Equation 1. X_{kj} is a (1×8) vector, representing the coding for "no preference", "\$0", "multi-word search", "partial-word search", "extra privacy", "maximum privacy", "5MB local storage" and "sync ability" for email service j in question k. β is the vector for part-worth utilities. The coding for "no preference" is 1 when "no preference" option is chosen by a participant, and 0 otherwise. We included this additional factor in order to get better model fitting and parameter estimation [95].

$$P_{nkj} = \frac{exp\,(X_{kj}\beta)}{\sum_{i=1}^{J} exp(X_{ki}\beta)} \tag{1}$$

Because part-worth values are scaled arbitrarily, the values themselves cannot be directly compared across features; instead, only the difference in utility in changing from one

option to another can be compared. For example, Table 5 shows that extra privacy has a part-worth utility of 0.190, while enabling partial-word search has a part-worth utility of 0.392. This cannot be interpreted directly to mean that extra privacy is of higher value than partial-word search. Instead, we see in Table 6 that the utility increase for enabling partial-word search is 0.785, while the increase from standard to extra privacy is 1.599. This means that when comparing these two option upgrades, the added privacy is more valuable.

For each part-worth utility, we also report a 95% confidence interval. The confidence interval helps to establish, with statistical confidence, the strength of apparent differences between features and options.

To more easily compare features, we use part-worth utilities to calculate the *relative importance* of each feature. Relative importance is defined as the ratio of the utility range for one feature to the sum of utility ranges across all features. Thus, relative importance (typically reported as a percentage) indicates how much a given feature contributes to a user's overall decision-making. In reporting aggregated relative importance, we calculate and then average the relative importance for all participants, rather than computing importance from average utilities.

Finally, to ease interpretation of the results, we calculate each utility change in terms of dollars. To do this, we divide the utility change from \$1.99 to free by the price difference (trivially, \$1.99); this yields the amount of utility change that is equivalent to \$1.99 per month. We then scale the utility changes for other features accordingly.

5.2.3. Estimated user preferences. Table 5 presents the aggregated part-worth utilities and relative importance for each feature. Considering each feature independently, our results show the expected relations: participants prefer the options that are inherently "better" (such as free rather than paid, and more privacy rather than less). This serves as an additional face validation of our results.

We find that with a relative importance of 32.6%, price is on average the most important factor influencing participants' choice of an email service. Privacy, at 24.2%, is second. Overall, participants were willing to pay \$0.86 per month to upgrade from standard to extra privacy, plus an additional \$0.55 to upgrade to maximum privacy. This indicates that the marginal value of privacy, as with many goods, is decreasing: our participants seem to believe that once some privacy improvement is made, further increases are less valuable. Finding that participants value privacy relatively cheaply aligns well with previous work [96], [97].

The third most important factor, after price and privacy, is local storage, with a relative importance of 17.4%. From Table 6, we see that reducing local storage from 500 to 5 MB is worth about \$0.68 per month. This is lower than, but somewhat comparable to, the \$0.86 value of a privacy upgrade from standard to extra, and higher than the value of an upgrade from extra to maximum privacy. This suggests that to our participants, local storage is almost as important as privacy improvements, likely because emails

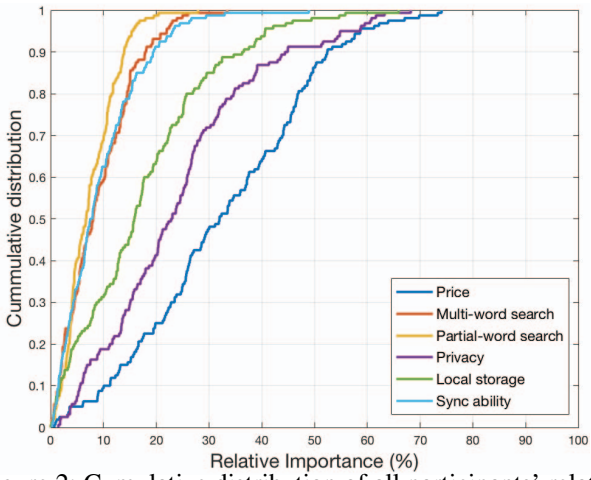


Figure 2: Cumulative distribution of all participants' relative importance for each feature.

are increasingly sent and received on mobile devices that have relatively limited storage.

The remaining features — synchronization, multi-word search, and partial-word search — each have a relative importance under 10%, with dollar values of \$0.50 or less. These values are somewhat comparable to the value of upgrading from extra to maximum privacy in our taxonomy. Further, upgrading any two of these three features is as or more valuable than increasing from standard to extra privacy. Overall, then, while our participants value privacy more than these other features, the added value is limited.

It is perhaps somewhat surprising that synchronization is rated with such low importance, given that emails are increasingly accessed on multiple devices; however, Cecchinato et al. report that email is managed differently on different devices, and in many cases separate devices manage different accounts [4]. They further report that searching email occurs primarily on laptops and PCs rather than on smartphones. This may help to explain the relatively low value placed on synchronization in our results.

The relative unimportance of complex search capabilities also aligns well with prior work. For example, a recent large-scale study found that the majority of search queries contain only one word [6]. Earlier work reported the existence of "many" partial-word queries [53], but we find no other evidence that they are frequently used. Several large webmail services (e.g., Gmail) support partial-word searches only in limited contexts, so it may be that users have not developed the habit of depending on this feature.

5.2.4. Variation across participants. We next investigate in more detail how perceived relative importance for each feature varied among all participants. In Figure 2, we present the cumulative distribution of all participants' perceived relative importance for each feature. This figure shows that, as discussed above, price is overall most important, followed by privacy and then local storage, while the lowest-importance three features cluster closely together. Price,

Model	Factors	Coef.	St. Dev.	p-value
Std. to Extra	Collection	0.288	0.073	<0.001*
Std. to Max	Web	0.357	0.177	0.045*
	Collection	0.341	0.147	0.022*
	Awareness	0.508	0.297	0.0887
	Control	-0.493	0.232	0.035*

TABLE 8: Regression results for utility changes in privacy options. We separately model change from standard to extra and from standard to maximum privacy (R^2 = 0.083 and 0.075 respectively). These results indicate the final model chosen via backward selection. Statistically significant factors with p<0.05 are marked as *.

privacy, and local storage also show more diversity of importance; the largest preferences reach over 60-70%, while the smallest are very close to 0. On the contrary, participants have more consensus on the lower perceived relative importance of multi-word search, partial-word search and synchronization.

5.2.5. Correlation among features. Finally, we consider whether any of the features we tested are strongly correlated: that is, whether any of them are strongly preferred together, or are treated as mutually exclusive. We show the correlation matrix of non-baseline variables in the Hierarchical Baysian analysis in Table 7. Most of the off-diagonal elements in the table are fairly close to 0, which implies that there are no two particular features are strongly preferred together or mutually exclusive by participants. The largest correlation is partial-word search ability with standard privacy (0.35). This positive coefficient suggests an association between valuing partial-word search more and valuing privacy less.

5.3. Why do web skills and reported privacy concern matter?

We next consider how participants' internet skills and generic privacy concerns affect their preferences for privacy within the email choices we present. To do this, we apply linear regression to model observed utility change as a function of a participant's scores on the web skills index as well as the three subscales of the IUIPC. We model change from standard to extra privacy, and then separately change from standard to maximum privacy.

To prevent possibly over-fitting our models, we applied the standard backward-elimination model selection process, until the Akaike Information Criterion (AIC) is minimized [87]. Our final regression results are shown in Table 8: we present which variables are selected, estimated coefficients and their standard deviations, and p-values.

The first model includes as a factor only the IUIPC Collection subscale. Malhotra et al. define Collection as degree of concern about "the amount of individual-specific data possessed by others relative to the value of benefits received" [82]. This model indicates that on average, for each one-point increase in concern as measured on this Likert subscale, participants value upgrading from standard to extra privacy an additional 0.288 of utility. This increase is significant.

The second model includes all the examined factors. The results indicate that greater web skills, as well as a higher Collection score, are significantly correlated with an increase in the utility of upgrading from standard to maximum privacy. On the other hand, participants with high scores on IUIPC Control — that is, a stronger belief that privacy is primarily a function of consumer control — place significantly lower value on upgrading from standard to maximum privacy. (The IUIPC Awareness score was not significantly correlated.)

Overall, these results suggest that people who are more tech-savvy, and who are more concerned about data collection, value privacy more highly; on the other hand, those who believe in the importance of individual control over data do not value it as strongly. We hypothesize that perhaps this is because the privacy options we offered to not exhibit individual fine-grained control, but rather allow the email service to observe information, or not.

6. Discussion

Our study takes the first step in characterizing user-facing tradeoffs inherent in supporting search for an encrypted communication solution. We consider six features that provide interesting separation across the space of possible approaches: price, multi-word search ability, partial-word search ability, privacy, local storage occupation, and synchronizing capability.

We found that among the non-monetary features we evaluated, privacy ranked highest overall, with local storage second. However, the value of a secondary privacy improvement — defined in our study as moving from extra to maximum privacy — was slightly less than the value of changing from 500 to 5 MB of local storage. Advanced search features using multiple words and partial-word matching were relatively less valuable overall.

These results have important implications for the design of encrypted email systems, and in particular for supporting search features. Current encrypted instant-messaging systems rely on a local-index solution, which provides maximum privacy (assuming no endpoint compromise) but also requires additional local storage. Searchable encryption approaches, in contrast, give up some degree of privacy, and limit the flexibility of the searches that can be conducted, but reduce the requirement for local storage.

Our results suggest, then, that in the email context local indexing may largely be suitable, but there may also be a niche for searchable encryption. For some users, the potential reduction in privacy relative to a local index will be worthwhile, in order to save on storage space. In particular, searchable encryption may have potential as a compromise: more privacy than the current standard, in which email services have plaintext access to all messages, while limiting some inconveniences that may prevent broader adoption of end-to-end encryption.

In line with prior work, we find that most users do not find significant value in advanced searching capabilities. Overall, this suggests that designers of searchable encryption schemes with email in mind concentrate on privacy and performance, rather than support for increasingly complex search operations.

Our results also show high variation in the relative values of privacy and storage among different participants. We also found that privacy concern and web-use skill are positively correlated with higher valuation of privacy, and particularly higher marginal valuation of additional privacy. This suggests that rather than attempting to develop one most suitable solution to enabling end-to-encryption for email, different solutions may be most appropriate for different user groups. Users with more web experience, with higher base levels of privacy concern, and/or with fewer limitations in e.g. local storage may prefer a local-index solution, while others may prefer a searchable-encryption solution. One possible option might be to provide a simple configuration switch offering the user more privacy/more storage or less privacy/less storage, and then implementing either a local index or searchable encryption accordingly. Simple tradeoff language may be appropriate for many users, potentially with a "more information" link to provide details to those who want or need them.

Prior work has shown that placing information about privacy up front can lead to users making more privacy-protective decisions [92]. Although the purpose of our study was not to design the best way to present information about email service privacy, our participants did value privacy highly when it was made an explicit decision feature. This suggests that highlighting privacy benefits and tradeoffs may help to increase the popularity of end-to-end-encrypted communications.

Future work. We explored only a small subset of possible tradeoffs inherent in supporting search for end-to-end encryption. Future work could apply similar approaches to test a broader array of features (for example, latency) or to test more options for each feature (a larger range of possible prices, more storage size options, or more detailed breakdowns of privacy models).

Further, we found in our pilot studies that the specifics of how feature information is worded affected participants' choices. While we believe that the descriptions we chose were reasonable for a first attempt to measure these tradeoffs, further work explicitly comparing how users respond to different option descriptions would be valuable. More generally, further work on how to explain threat models, privacy tradeoffs, and the guarantees that encryption does and does not provide are urgently needed to ensure that users do not operate with a false sense of security [3].

Finally, we know that expressed preferences, even in a conjoint analysis study designed to operationalize relative value, do not always match real-world behavior. As such, it would be useful to design a follow-up field study, in which participants live with the consequences of the tradeoffs they select for days or weeks, to see how this changes their opinions.

Acknowledgments

We thank our anonymous reviewers and shepherd for their thoughtful and constructive comments and suggestions. This research was supported in part by the Maryland Procurement Office under contract no. 00024010.

References

[1] WhatsApp, https://www.whatsapp.com/.
[2] Apple, "Learn how to use messages," "https://support.apple.com/explore/messages".
[3] R. Abu-Salma, M. A. Sasse, J. Bonneau, A. Danilova, A. Naiakshina, and M. Smith, "Obstacles to the adoption of secure communication tools," in *Proc. of IEEE S&P '17*, May 2017.
[4] M. E. Cecchinato, A. Sellen, M. Shokouhi, and G. Smyth, "Finding email in a multi-account, multi-device world," in *Proc. of CHI '16*. ACM, 2016, pp. 1200–1210.
[5] Q. Ai, S. T. Dumais, N. Craswell, and D. Liebling, "Characterizing email search using large-scale behavioral logs and surveys," in *Proc. of WWW '17*, 2017, pp. 1511–1520.
[6] D. Carmel, L. Lewin-Eytan, A. Libov, Y. Maarek, and A. Raviv, "The demographics of mail search and their application to query suggestion," in *Proc. of WWW '17*, 2017, pp. 1541–1549.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. of IEEE S&P '00*, 2000, pp. 44–55.
[8] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in *Proc. of CCS '06*. ACM, 2006, pp. 79–88.
[9] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in *Proc. of CCS '12*. ACM, 2012, pp. 965–976.
[10] S. Kamara and C. Papamanthou, *Parallel and Dynamic Searchable Symmetric Encryption*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 258–274.
[11] E. Stefanov, C. Papamanthou, and E. Shi, "Practical dynamic searchable encryption with small leakage," in *NDSS '14*. Internet Society, 2014.
[12] R. Bost, "$\Sigma o \phi o \varsigma$: Forward secure searchable encryption," in *Proc. of CCS '16*, 2016, pp. 1143–1154.
[13] P. E. Green and V. Srinivasan, "Conjoint analysis in marketing: New developments with implications for research and practice," *Journal of Marketing*, vol. 54, no. 4, pp. 3–19, 1990.
[14] H. Sattler and A. Hartmann, *Commercial Use of Conjoint Analysis*. Wiesbaden: Gabler, 2008, pp. 103–119.
[15] M. Vriens, "Solving marketing problems with conjoint analysis," *Journal of Marketing Management*, vol. 10, no. 1-3, pp. 37–55, 1994.
[16] Y. Pu and J. Grossklags, "Towards a model on the factors influencing social app users' valuation of interdependent privacy," in *Proc. of PETS '16*, 2016, pp. 61–81.
[17] H. Krasnova, T. Hildebrand, and O. Guenther, "Investigating the value of privacy in online social networks: Conjoint analysis," in *ICIS '09 Proc.*, 2009.
[18] D. Burda and F. Teuteberg, "Understanding the benefit structure of cloud storage as a means of personal archiving - a choice-based conjoint analysis," in *ECIS '14*, 2014.
[19] P. R. Zimmermann, *The Official PGP User's Guide*. Cambridge, MA, USA: MIT Press, 1995.
[20] GPGTools, "GPGTools," https://gpgtools.org/.
[21] "Mailvelope," https://www.mailvelope.com/en/.
[22] "Signal," https://whispersystems.org/.
[23] N. Unger, S. Dechand, J. Bonneau, S. Fahl, H. Perl, I. Goldberg, and M. Smith, "SoK: Secure messaging," in *Proc. of IEEE S&P '15'*, May 2015, pp. 232–249.
[24] A. Whitten and J. D. Tygar, "Why johnny can't encrypt: A usability evaluation of pgp 5.0," in *Proc. of SSYM'99*, Berkeley, CA, USA, 1999, pp. 14–14.

[25] S. Sheng, L. Broderick, J. J. Hyland, and C. A. Koranda, "Why johnny still can't encrypt: Evaluating the usability of email encryption software," in *Proc. of SOUPS '06*, 2006.

[26] S. Ruoti, J. Andersen, T. Hendershot, D. Zappala, and K. Seamons, "Private webmail 2.0: Simple and easy-to-use secure email," in *Proc. of UIST '16*. ACM, 2016, pp. 461–472.

[27] W. Tong, S. Gold, S. Gichohi, M. Roman, and J. Frankle, "Why King George III can encrypt," http://randomwalker.info/teaching/spring-2014-privacy-technologies/king-george-iii-encrypt.pdf, 2014.

[28] J. Lausch, O. Wiese, and V. Roth, "What is a secure email," in *Proc. of EuroUSEC '17*. Internet Society, 2017.

[29] S. Ruoti, N. Kim, B. Burgon, T. van der Horst, and K. Seamons, "Confused johnny: When automatic encryption leads to confusion and mistakes," in *Proc. of SOUPS '13*. ACM, 2013, pp. 5:1–5:12.

[30] S. Ruoti, J. Andersen, S. Heidbrink, M. O'Neill, E. Vaziripour, J. Wu, D. Zappala, and K. Seamons, "'We're on the same page': A usability study of secure email using pairs of novice users," in *Proc. of CHI '16*. ACM, 2016, pp. 4298–4308.

[31] S. L. Garfinkel and R. C. Miller, "Johnny 2: A user test of key continuity management with s/mime and outlook express," in *Proc. of SOUPS '05*, 2005, pp. 13–24.

[32] S. Fahl, M. Harbach, T. Muders, and M. Smith, "Confidentiality as a service – usable security for the cloud," in *TrustCom '12*, June 2012, pp. 153–162.

[33] M. D. Ryan, "Enhanced certificate transparency and end-to-end encrypted mail," in *NDSS '14*. Internet Society, 2014.

[34] M. S. Melara, A. Blankstein, J. Bonneau, E. W. Felten, and M. J. Freedman, "Coniks: Bringing key transparency to end users," in *USENIX Security '15*, 2015, pp. 383–398.

[35] S. Fahl, M. Harbach, T. Muders, M. Smith, and U. Sander, "Helping johnny 2.0 to encrypt his facebook conversations," in *Proc. of SOUPS '12*. ACM, 2012, pp. 11:1–11:17.

[36] W. Bai, M. Namara, Y. Qian, P. G. Kelley, M. L. Mazurek, and D. Kim, "An inconvenient trust: User attitudes toward security and usability tradeoffs for key-directory encryption systems," in *Proc. of SOUPS '16*, Denver, CO, 2016, pp. 113–130.

[37] A. Lerner, E. Zeng, and F. Roesner, "Confidante: Usable encrypted email: A case study with lawyers and journalists," in *Proc. of EuroS&P '17'*, April 2017, pp. 385–400.

[38] M. Shirvanian and N. Saxena, "On the security and usability of crypto phones," in *Proc. of ACSAC '15*, 2015, pp. 21–30.

[39] J. Tan, L. Bauer, J. Bonneau, L. F. Cranor, J. Thomas, and B. Ur, "Can unicorns help users compare crypto key fingerprints?" in *Proc. of CHI '17*. ACM, 2017, pp. 3787–3798.

[40] D. J. Solove, "I've got nothing to hide and other misunderstandings of privacy," *San Diego Law Review*, vol. 44, no. 289, p. 745, 2007, associate Professor, George Washington University Law School; J.D., Yale Law School.

[41] S. Gaw, E. W. Felten, and P. Fernandez-Kelly, "Secrecy, flagging, and paranoia: Adoption criteria in encrypted email," in *Proc. of CHI '06*. ACM, 2006, pp. 591–600.

[42] A. D. Luca, S. Das, M. Ortlieb, I. Ion, and B. Laurie, "Expert and non-expert attitudes towards (secure) instant messaging," in *Proc. of SOUPS '16*, Denver, CO, 2016, pp. 147–157.

[43] S. E. McGregor, P. Charters, T. Holliday, and F. Roesner, "Investigating the computer security practices and needs of journalists," in *USENIX Security '15*, Washington, D.C., 2015, pp. 399–414.

[44] B. Lau, S. Chung, C. Song, Y. Jang, W. Lee, and A. Boldyreva, "Mimesis aegis: A mimicry privacy shield–a system's approach to data privacy on public cloud," in *USENIX Security '14*, 2014, pp. 33–48.

[45] D. Cash, J. Jaeger, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rosu, and M. Steiner, "Dynamic searchable encryption in very-large databases: Data structures and implementation," Cryptology ePrint Archive, Report 2014/853, 2014.

[46] P. Grofig, M. Haerterich, I. Hang, F. Kerschbaum, M. Kohler, A. Schaad, A. Schroepfer, and W. Tighzert, "Experiences and observations on the industrial implementation of a system to search over outsourced encrypted data," http://subs.emis.de/LNI/Proceedings/Proceedings228/article7.html, 2014.

[47] Bitglass, "http://www.bitglass.com/".

[48] Google, "https://github.com/google/encrypted-bigquery-client".

[49] Microsoft, "Always encrypted (database engine)," "https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine".

[50] "Skyhigh Networks," "https://www.skyhighnetworks.com/".

[51] B. Fuller, M. Varia, A. Yerukhimovich, E. Shen, A. Hamlin, V. Gadepally, R. Shay, J. D. Mitchell, and R. K. Cunningham, "Sok: Cryptographically protected database search," in *IEEE S&P '17*, May 2017, pp. 172–191.

[52] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins, "Stuff I've seen: A system for personal information retrieval and re-use," in *Proc. of SIGIR '03*. ACM, 2003, pp. 72–79.

[53] M. Harvey and D. Elsweiler, *Exploring Query Patterns in Email Search*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 25–36.

[54] S. Whittaker, T. Matthews, J. Cerruti, H. Badenes, and J. Tang, "Am i wasting my time organizing email?: A study of email refinding," in *Proc. of CHI '11*. ACM, 2011, pp. 3449–3458.

[55] J. Jordan, "53% of emails opened on mobile; outlook opens decrease 33%," https://litmus.com/blog/53-of-emails-opened-on-mobile-outlook-opens-decrease-33, Jan. 2015.

[56] Adestra, "Consumer adoption and usage study," http://www.adestra.com/resources/downloadable-reports/consumer-adoption-and-usage-study/, 2016.

[57] Y. Pu and J. Grossklags, "Valuating friends' privacy: Does anonymity of sharing personal data matter?" in *Proc. of SOUPS '17)*, 2017, pp. 339–355.

[58] T. Moataz and A. Shikfa, "Boolean symmetric searchable encryption," in *Proc. of ASIA CCS '13*. ACM, 2013, pp. 265–276.

[59] E. Chen, I. Gomes, B. Saavedra, and J. Yucra, "Cocoon: encrypted substring search," 2015, https://courses.csail.mit.edu/6.857/2016/files/29.pdf.

[60] C. Melissa and S. Emily, "Substring-searchable symmetric encryption," *Proc. on Privacy Enhancing Technologies*, vol. 2015, p. 263, Apr. 2015.

[61] X. Meng, S. Kamara, K. Nissim, and G. Kollios, "Grecs: Graph encryption for approximate shortest distance queries," in *Proc. of CCS '15*. ACM, 2015, pp. 504–517.

[62] T. Moataz and E.-O. Blass, "Oblivious substring search with updates," *IACR Cryptology ePrint Archive*, vol. 2015, p. 722, 2015.

[63] Y. Ishai, E. Kushilevitz, S. Lu, and R. Ostrovsky, *Private Large-Scale Databases with Distributed Searchable Symmetric Encryption*. Cham: Springer International Publishing, 2016, pp. 90–107.

[64] A. Oulasvirta and L. Sumari, "Mobile kits and laptop trays: Managing multiple devices in mobile information work," in *Proc. of CHI '07*. ACM, 2007, pp. 1127–1136.

[65] "ProtonMail," https://protonmail.com/.

[66] "Threema," https://threema.ch/en.

[67] Google, "How gmail ads work," https://support.google.com/mail/answer/6603?hl=en, 2017.

[68] V. Pappas, F. Krell, B. Vo, V. Kolesnikov, T. Malkin, S. G. Choi, W. George, A. Keromytis, and S. Bellovin, "Blind seer: A scalable private dbms," in *Proc. of IEEE S&P '14*, 2014, pp. 359–374.

[69] B. A. Fisch, B. Vo, F. Krell, A. Kumarasubramanian, V. Kolesnikov, T. Malkin, and S. M. Bellovin, "Malicious-client security in blind seer: A scalable private dbms," in *Proc. of IEEE S&P '15*, May 2015, pp. 395–410.

[70] D. Pouliot and C. V. Wright, "The shadow nemesis: Inference attacks on efficiently deployable, efficiently searchable encryption," in *Proc. of CCS '16*, 2016, pp. 1341–1352.

[71] Y. Zhang, J. Katz, and C. Papamanthou, "All your queries are belong to us: The power of file-injection attacks on searchable encryption," in *USENIX Security '16*, Austin, TX, 2016, pp. 707–720.

[72] M. Naveed, "The fallacy of composition of oblivious ram and searchable encryption," Cryptology ePrint Archive, Report 2015/668, 2015, http://eprint.iacr.org/2015/668.

[73] P. E. Green and V. R. Rao, "Conjoint measurement for quantifying judgmental data," *Journal of Marketing Research*, vol. 8, no. 3, pp.

355–363, 1971.
[74] P. E. Green and V. Srinivasan, “Conjoint analysis in consumer research: Issues and outlook,” *Journal of Consumer Research*, vol. 5, no. 2, pp. 103–123, 1978.
[75] W. S. Desarbo, V. Ramaswamy, and S. H. Cohen, “Market segmentation with choice-based conjoint analysis,” *Marketing Letters*, vol. 6, no. 2, pp. 137–147, 1995.
[76] G. Box, J. Hunter, and W. Hunter, *Statistics for experimenters: design, innovation, and discovery*, ser. Wiley series in probability and statistics. Wiley-Interscience, 2005.
[77] M. Bech, T. Kjaer, and J. Lauridsen, “Does the number of choice sets matter? results from a web survey applying a discrete choice experiment,” *Health Economics*, vol. 20, no. 3, pp. 273–286, 2011.
[78] H. Aizaki and K. Nishimura, “Design and analysis of choice experiments using r: A brief introduction,” *Agricultural Information Research*, vol. 17, no. 2, pp. 86–94, 2008.
[79] “Amazon mechanical turk,” https://www.mturk.com/mturk/welcome.
[80] E. Peer, J. Vosgerau, and A. Acquisti, “Reputation as a sufficient condition for data quality on amazon mechanical turk,” *Behavior Research Methods*, vol. 46, no. 4, pp. 1023–1031, 2014.
[81] E. Hargittai and Y. P. Hsieh, “Succinct survey measures of web-use skills,” *Soc. Sci. Comput. Rev.*, vol. 30, no. 1, pp. 95–107, Feb. 2012.
[82] N. K. Malhotra, S. S. Kim, and J. Agarwal, “Internet users’ information privacy concerns (iuipc): The construct, the scale, and a causal model,” *Info. Sys. Research*, vol. 15, no. 4, pp. 336–355, Dec. 2004.
[83] S. Preibusch, “Guide to measuring privacy concern: Review of survey and observational instruments,” *Int. J. Hum.-Comput. Stud.*, vol. 71, no. 12, pp. 1133–1143, Dec. 2013.
[84] P. E. Rossi and G. M. Allenby, “Bayesian statistics and marketing,” *Marketing Science*, vol. 22, no. 3, pp. 304–328, Sep. 2003.
[85] P. Rossi, *bayesm: Bayesian Inference for Marketing/Micro-Econometrics*, https://CRAN.R-project.org/package=bayesm, 2015, r package version 3.0-2.
[86] M. Bech and D. Gyrd-Hansen, “Effects coding in discrete choice experiments,” *Health Economics*, vol. 14, no. 10, pp. 1079–1083, 2005.
[87] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society, series B*, vol. 68, pp. 49–67, 2006.
[88] R. Kang, S. Brown, L. Dabbish, and S. Kiesler, “Privacy attitudes of mechanical turk workers and the u.s. public,” in *Proc. of SOUPS ’14*, 2014, pp. 37–49.
[89] R. Vetschera and G. Kainz, “Do self-reported strategies match actual behavior in a social preference experiment?” *Group Decision and Negotiation*, vol. 22, no. 5, pp. 823–849, 2013.
[90] E. Hargittai and A. Marwick, ““what can i really do?” explaining the privacy paradox with online apathy,” *International Journal of Communication*, vol. 10, no. 0, 2016.
[91] A. Nuno and F. A. St. John, “How to ask sensitive questions in conservation: A review of specialized questioning techniques,” *Biological Conservation*, vol. 189, pp. 5–15, 2015.
[92] P. G. Kelley, L. F. Cranor, and N. Sadeh, “Privacy as part of the app decision-making process,” in *Proc. of CHI ’13*. ACM, 2013, pp. 3393–3402.
[93] S. Barge and H. Gehlbach, “Using the theory of satisficing to evaluate the quality of survey data,” *Research in Higher Education*, vol. 53, no. 2, pp. 182–200, Mar 2012.
[94] K. Train, *Discrete Choice Methods with Simulation*. SUNY-Oswego, Department of Economics, 2003.
[95] R. Haaijer, W. Kamakura, and M. Wedel, “The ‘no-choice’ alternative in conjoint choice experiments,” *International Journal of Market Research*, vol. 43, 2001.
[96] A. Acquisti and J. Grossklags, “An online survey experiment on ambiguity and privacy,” *Communications & Strategies*, vol. 88, pp. 19–39, 2012.
[97] J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti, “The effect of online privacy information on purchasing behavior: An experimental study,” in *ICIS ’07 Proc.*, 2007.

Appendix

This appendix contains the full 16 email service choice questions in our research. During the survey, the sequence of questions was randomized for each participant.