Westminster Papers in Communication and Culture

Musiani, F. and Ermoshina, K. (2017). What is a *Good* Secure Messaging Tool? The EFF Secure Messaging Scorecard and the Shaping of Digital (Usable) Security. *Westminster Papers in Communication and Culture,* 12(3), 51–71, DOI: https://doi.org/10.16997/wpcc.265

RESEARCH ARTICLE

What is a *Good* Secure Messaging Tool? The EFF Secure Messaging Scorecard and the Shaping of Digital (Usable) Security

Francesca Musiani and Ksenia Ermoshina

Institute for Communication Sciences – ISCC, CNRS/Paris-Sorbonne/UPMC, FR

Corresponding author: Francesca Musiani (francesca.musiani@cnrs.fr)

In today's diverse and crowded landscape of messaging systems, what are the most secure and usable tools? The Electronic Frontier Foundation (EFF), a digital rights group based in San Francisco, CA, has been considering this question for a long time. Their most prominent initiative in this regard has been the 2014 release of the Secure Messaging Scorecard (SMS), a 7-criteria evaluation of 'usable security' in messaging systems. While the 2014 version of the SMS (now 1.0) displays a number of apparently straightforward criteria, a first look into the backstage shows that the selection and formulation of these criteria has been anything but linear, something that has been made particularly evident by the EFF's recent move to renew and update the SMS. Indeed, in a digital world where the words security and privacy are constantly mobilized with several different meanings, it seems relevant to analyse the SMS's first release, and the subsequent discussions and renegotiations of it, as processes that co-produce particular definitions of security, of defence against surveillance, and of privacy protection. This article argues that, by means of the SMS negotiations around the categories that are meaningful to qualify and define encryption, the EFF is in fact contributing to the shaping of what makes a 'good' secure messaging application, and what constitutes a 'good' categorization system to assess (usable) security, able to take into account all the 'relevant' aspects – not only technical but social and economic.

Keywords: Science and technology studies; categories; Internet governance; encryption; secure messaging; Electronic Frontier Foundation

1. Introduction

In today's diverse and crowded landscape of messaging systems, what are the most secure and usable tools? The Electronic Frontier Foundation (EFF), a digital rights group based in San Francisco, has been considering this question for a long time. Their most prominent initiative in this regard has been the 2014 release of the Secure Messaging Scorecard (SMS),[1] a 7-criteria evaluation of 'usable security' in messaging systems.

[1] https://www.eff.org/node/82654.

While the 2014 version of the SMS (now 1.0) displays a number of apparently straightforward criteria – including, but not limited to, encryption of data in transit, ability to verify contacts' identities, available documentation for security design assessment and whether a code audit has happened in the recent past – a first look into the backstage shows that the selection and formulation of these criteria has been anything but linear.[2] This has been made particularly evident by the EFF's recent move to update the SMS; acknowledging that 'Though all of those criteria are necessary for a tool to be secure, they can't guarantee it; security is hard, and some aspects of it are hard to measure', the foundation proceeds to announce that 'That's why we're working on a new, updated, more nuanced format for the Secure Messaging Guide'.[3]

Indeed, in a digital world where the words security and privacy are constantly mobilized with several different meanings – even within the same debates and by a priori alike actors – it seems relevant to analyse the SMS's first release, and the subsequent discussions and renegotiations of it, as processes that de-stabilize, negotiate and possibly re-stabilize particular definitions of security, of defence against surveillance, and of privacy protection. This article argues that, by means of the SMS negotiations around the categories that are meaningful to qualify and define encryption, the EFF is in fact contributing to shape what makes a 'good' secure messaging application and what constitutes a 'good' measurement system to assess (usable) security, able to take into account all the 'relevant' aspects – not only technical but social and economic.

The article draws primarily from original qualitative interviews conducted with members of the EFF and on the examination of the first version of the SMS, as well as other documentary sources. From a theoretical standpoint, the article draws upon an important body of work in the field of science and technology studies (STS) that has addressed the 'making of' systems of classification, categorization and measurement as a crucial component of human interaction and governance processes (notably Bowker and Star, 1999), linking it to previous and current work from both authors on the governance of networked systems and secure messaging tools (Musiani et al., 2016; Ermoshina, Musiani & Halpin, 2016).

2. 'Sorting things out', a long-term STS preoccupation

As Geoffrey Bowker and Susan Leigh Star remind us in their seminal work *Sorting Things Out* (1999), issues such as the origin of categorization and classification systems and the ways in which they shape the boundaries of the communities that use them, have been an important preoccupation of social sciences in the last century. Scholars in the field of Science and Technology Studies (STS) in particular have explored these systems as tools that shape the environments or the infrastructures they seek to categorize, and addressed their particular status as both a 'thing and an action', having both material and symbolic dimensions (Bowker and Star, 1999: 285–6). As our empirical contribution in this paper will show, the EFF's attempt to define an appropriate categorization system to assess the quality of secure messaging tools is indeed thing and action at once, co-shaping the world it seeks to organize.

2.1. Classification, categorization, and the shaping of communities of practice

From an STS perspective, classification and categorization processes are strictly linked to the shared perception different actors are able to have of themselves as belonging to a community. In many cases, they highlight the boundaries that exist between communities and constitute the terrain where they might get closer – or drift further apart:

[2] E.g. the discussion of the code audit criterion at Peter Eckersley, 'What Makes a Good Security Audit?', EFF Deeplinks, 8 November 2014, https://www.eff.org/deeplinks/2014/11/what-makes-good-security-audit, which will be discussed in more detail at 3.1.

[3] https://www.eff.org/secure-messaging-scorecard.

> 'Information technologies used to communicate across the boundaries of disparate communities [...] These systems are always heterogeneous. Their ecology encompasses the formal and the informal, and the arrangements that are made to meet the needs of heterogeneous communities—some cooperative and some coercive' (ibid, 286).

Categorization processes, as Goodwin (1996: 65) reminds us, are meant to 'establish [...] an orientation towards the world', to construct shared meanings in larger organizational systems.

Borrowing from Cole (1996: 117), Bowker and Star point out that the categories produced by such processes are both conceptual (as they are resources for organizing abstractions, returning patterns of action and change) and material (because they are inscribed, affixed to material artefacts). The act of using any kind of representation, of schematization through simplification, is a complex achievement, an 'everyday, [yet] impossible action' (Bowker and Star, 1999: 294) that is however necessary to become part of a 'community of practice' (Lave and Wenger, 1991), or in Becker's words, it is a set of relations among people doing things together (Becker, 1986). The community structure is constituted by 'routines' and 'exceptions' as identified by the categorization system – the more the shared meaning of this system is stabilized among the members of the community, the more the community itself is stabilized as such:

> Membership in a community of practice has as its sine qua non an increasing familiarity with the categories that apply to all of these. As the familiarity deepens, so does one's perception of the object as strange or of the category itself as something new and different (Bowker and Star, 1999: 294).

We will see how, in the highly unstable environment that is the EFF's pioneering attempt to categorize secure messaging tools with a view to providing guidance on their quality, the embryo of a community of practice that goes beyond the relatively homogeneous group of cryptography developers starts to be born. It includes users of different expertise, trainers and civil liberties activists, and simultaneously reveals the manifold points of friction between these groups.

Classifications and categorizations, Bowker and Star later point out (ibid, 319) are also 'powerful technologies', whose architecture is simultaneously informatic and moral. Due to their embeddedness in working infrastructures, they can become relatively invisible as they progressively stabilize, without losing their power. Thus, categorization systems should be acknowledged as a significant site of political, ethical and cultural work – three aspects that our analysis of the SMS negotiations will unfold. In these three respects, categories are performative (Callon, 2009): the reality of the everyday practices they subtend reveals that, far from being 'enshrined [...] in procedures and stabilized conventional principles that one merely needs to follow in order to succeed' (Denis, 2006: 12, our translation), they actively participate in the construction of the relation between the different actors that have a stake, or a role, in the categorized environment; categories, in this view, are one of the components of a complex network of actors and technologies.

2.2. Secure messaging tools: A field in the making

Before we move on to analysing the categorization system this paper is about – and what we believe it brings to the understanding of this special issue's theme of 'rethinking privacy' – it is useful to give some elements of context about the ensemble of tools, and more broadly the field, it seeks to address. We will not engage here in a literature review about encryption, security and privacy, for two reasons. First, because volumes have been and are being written on each of these topics, and it would be a daunting task to even start acknowledging them with any degree of completeness. Second, because the aim of this paper is not to assess the validity of the arguments supporting different scholarly appreciations of each topic and position ourselves among them. Instead, we wish to show how understandings of 'good' encryption, security and privacy emerge – more often than not, in a non-academic and bottom-up fashion – from a specific categorization attempt that, because of its pioneering role, raises a number of questions, tensions and reflections among and across the interested actors. That being said, this section will briefly address the current state of the field of secure messaging tools.[4]

Whilst it of course predates the Snowden revelations – but, certainly, was further spurred by them – secure messaging is a lively and constantly evolving ecosystem of standardized and non-standardized projects. Developers seek, in particular, to apply the technique of end-to-end encryption[5] to messaging systems: among the most widely-known tools pertaining to this category are Signal, Telegram and WhatsApp. In terms of underlying protocols, the field converges towards the Signal protocol, with a few attempts to decentralize/federate messaging apps that do not, however, count many users beyond the technical community. Despite the prevalence of free and open source software projects, proprietary software is not absent in this landscape. Open source itself is multi-layered and sometimes hybrid, with the code on the client side being open source and the server side being proprietary (e.g. in Telegram). Perhaps unsurprisingly, the proprietary features are more important in applications designed for business-to-business use, while free and open source software is predominant for tools designed for activists and tech-savvy users. This transparency of code and encryption protocols is, in most cases, aimed not only at improving the project, but also at creating an emulation around the project producing a collective reflection of experts, beta-testers and advanced users.

According to a post-Snowden systematization of knowledge paper on secure messaging, the field currently suffers from the 'lack of a clear winner in the race for widespread deployment and the persistence of many lingering unsolved research problems', as well as discrepancies between 'grandiose claims' and actual provided security – an issue that tools such as the SMS have attempted to tackle (Unger et al., 2015). Part of the reason for the field's diversity and complexity is the relatively short life span of several projects. While in more than a few cases, the motives behind this are primarily related to a technical experimentation that did not deliver as hoped or expected, a number of additional factors may also be responsible, including the failure to develop an economic model, the internal governance of FOSS development groups, and the inability to rally a critical mass of users around the app (possibly due to a lack of ease-of-use). The target audience of the applications, especially those born post-Snowden, is far from being limited to tech-savvy and activist groups; several projects are aimed at widespread use. A majority of members of the technical crypto community consider user-friendliness and usability as the main issue that stands between the wish for large-scale adoption and its realization in practice, although this take has been challenged by scholars as a 'forced responsabilisation' of users to the detriment of the development of resilient collective digital security strategies (Kazansky, 2015) and as a 'delegation' of technical matters

[4] Borrowing from the conclusions of a survey conducted within the first year of the NEXTLEAP project, partly summarized in (Ermoshina et al., 2016) and shortly publicly available as Deliverable 3.1 of the NEXTLEAP project on the dedicated page of the CORDIS website http://cordis.europa.eu/project/rcn/199878_en.html.

[5] Communication system where only the communicating users can read the messages, thanks to encryption keys only known to them.

to 'progressive techies' despite a widespread societal desire to develop technologies for social justice (Aouragh et al., 2015).

End-to-end encrypted messaging tools are currently at the centre of a powerful 'double' narrative. If on one hand the discourse on empowerment and better protection of fundamental civil liberties is very strong, several projects show in parallel a desire, or a need, to defend themselves from allegations of links to terrorism (Sanger and Perlroth, 2015). This latter narrative is fuelled by previous and current ones about decentralized technologies and peer-to-peer, with their history of allegedly 'empowering-yet-illegal' tools. These issues are taking place in the broader context of discussions about civil liberties and governance by infrastructure (Musiani et al., 2016), some of them particularly related to encryption (or the breaking of it), such as the Apple vs. FBI case and WhatsApp proposing, since April 2016, encryption by default. Indeed, after the Snowden revelations, several companies, in particular those based in the United States, have implemented a number of cryptography-based organizational and technical responses aimed at restoring user trust in their cloud-based services. This dynamic has been identified as a 'cryptographic turn' opening up new issues and questions from both legal and political standpoints (Rubinstein & Van Hoboken, 2014; Gürses, Kundnani & Van Hoboken, 2016) and is considered a new phase of the 1990s 'Crypto Wars' (Froomkin & McLaughlin, 2016).

2.3. Methodology and ethical guidelines

This paper is supported by a qualitative methodology and draws from the research conducted within the H2020 project NEXTLEAP[6] on secure messaging ecosystems, encryption and decentralization. We have been conducting fieldwork since January 2016, including observations and in-depth interviews. We have so far interviewed 53 developers, users and digital security trainers from various countries (low-risk Western countries and higher risk contexts such as Ukraine, Russia, Iran, Lebanon, Kenya, and Egypt). We have interviewed the most popular messaging apps teams: from centralized projects such as Signal or Wire to Conversations, Briar, Ricochet, LEAP/Pixelated, ChatSecure and other decentralized applications.[7] The choices of apps to interview were based on a preliminary mapping of 30 case studies (Ermoshina, Musiani & Halpin, 2016). Our interviews with developers were semi-structured and lasted from 1 hour to 3 hours long. We also conducted web ethnography and analysed the GitHub/GitLab pages of the studied projects, as well as user feedback and dedicated press publications.

As for the present paper focusing on the Electronic Frontier Foundation, we started our research by consulting the appropriate documents, which will be extensively cited (Version 1.0 of the SMS, online discussions and blog posts discussing it). Once the 'check back soon' notice was made public by the EFF (see **Figure 2**), it was a first-rate invitation for STS researchers like us, to investigate the backstage of current negotiations, which could only be achieved via original, in-depth interviews with the EFF team. Three interviews, between 1 hour and 2 hours and a half in length, were conducted with the person in charge of the first SMS (R1, 20 December 2016), the coordinator of the second SMS (R2, 9 December 2016), and the trainer and coordinator of the EFF Surveillance Self-Defense Guide (R3, 30 November 2016). Additionally, the analysis in this paper is supported by several other interviews with cryptoparty organizers in Austria and Germany (November and December 2016) and informational security trainers in Ukraine (January 2017) about their usages of SMS 1.0 during training sessions and two

[6] http://nextleap.eu.

[7] The full list of projects, methodology of interviews (including questionnaires), as well as key findings from these interviews are presented in a dedicated paper (Ermoshina, Halpin & Musiani, 2017).

interviews with the CTO and chief of marketing of Wire, a secure messaging app to be added in the second version of the SMS. Hence, it was particularly relevant for us to interview the Wire team on their interactions with the EFF in order to understand the making of the second version of SMS and how SMS influences (or not) the technical and legal work of a secure messaging application development team.

This article focuses on the negotiations around the technical properties selected as part of the categorization, seen primarily through the viewpoint of EFF members. In doing so, its primary focus is not on the Electronic Frontier Foundation as a 'political economy actor', although it should be acknowledged that the relationships of this non-profit organization vis-à-vis other similar organizations, the private sector, political institutions in the United States and the 'crypto' technical community have contributed to the selection of the particular set of encrypted secure messaging applications that made it to the SMS, and to what could be considered as an 'over-inclusion' of US-based companies, perhaps to the detriment of apps from other regions or language spaces.

As for ethical guidelines, a specific protocol was developed in order to protect the privacy of our respondents. We let users and developers suggest a secure tool of communication of their choice to us if they wished to do the interview online. The interview was recorded with an audio recorder isolated from the Internet. We used a dedicated encrypted hard drive to store the interviews. Before the interview we asked our respondents to carefully read two user-consent forms related to the study and ask all the questions regarding their privacy, their rights and our methodology. The two forms were written in collaboration with UCL usability researchers and based on the European General Data Protection Regulation. The documents included an Information Sheet and an Informed Consent Form. The first document explained the purpose of the interview, described the research project and clearly mentioned the sources of funding for the project; provided information on the length of the interview, but also information about the researcher, including her email, full name, academic affiliation and the address of the research institution. The second form described the procedures regarding data processing methods, the period and conditions of data storage; it emphasized the right of the interviewees to demand, at any moment, to withdraw their data from the research. A copy of each document was given to the interviewee. Different forms were used for users, trainers and developers. We respected the right of our interviewees to refuse answering a specific question and choose the degree of anonymization.[8]

3. The Secure Messaging Scorecard 1.0: Unveiling 'actual security'?

In November 2014, the Electronic Frontier Foundation released its Secure Messaging Scorecard. The SMS was announced as the first step of an awareness campaign aimed at both companies and users – a tool that, while abstaining from formal endorsement of particular products, aimed at providing guidance and reliable indications that 'projects are on the right track', in an increasingly complex landscape of self-labelled 'secure messaging products', in providing 'actual [...] security'.[9] According to R1, *'We tried to hit both audiences, we wanted to provide information for users and we also wanted to encourage tools to adopt more encryption to do things like release source code or undergo an audit'.*

[8] More detailed explanations of our ethical guidelines are available in (Ermoshina, Halpin & Musiani, 2017).

[9] https://www.eff.org/node/82654. This webpage, introducing SMS v1, is now preserved 'for purely historical reasons' on the EFF website. Citations in 3.1. are from this page unless otherwise noted.

3.1. Overcoming the security vs. usability trade-off

The EFF closely links the initiative to the Snowden revelations, mentioning that privacy and security experts have repeatedly called on the public to adopt widespread encryption in recent years; however, Snowden's whistle-blowing on governments' *'grabbing up communications transmitted in the clear'* has made widespread routine adoption of encrypting tools a matter of pressing urgency. R1 suggests that the need for such a tool came out of a need expressed by the general public:

> It kind of came out of conversations at EFF [...we] got a lot of queries with people asking which messaging tools they should use. So there have been a couple of projects in the past, like the guide 'Who has your back', a project that is a sort of scorecard of Terms of Service and how companies handle data, that's where the idea came from... try to use the same approach to put information out there that we thought was useful about different messaging tools.

In the EFF's view, adoption of encryption to such a wide extent 'boils down to two things: security and usability'. It is necessary that both things go hand in hand, while in most instances, the EFF observes a trade-off between the two:

> Most of the tools that are easy for the general public to use don't rely on security best practices–including end-to-end encryption and open source code. Messaging tools that are really secure often aren't easy to use; everyday users may have trouble installing the technology, verifying its authenticity, setting up an account, or may accidentally use it in ways that expose their communications.[10]

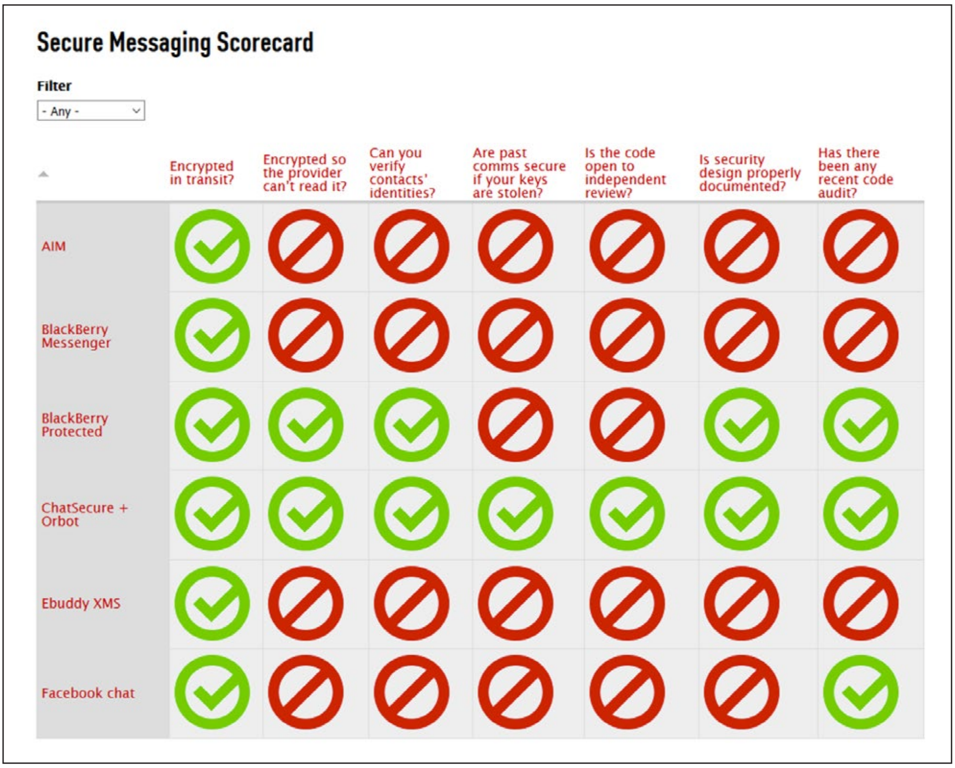
Citing collaborations with notable civil liberties organizations such as ProPublica and the Princeton University Research Center on Information and Technology Policy, the EFF presents the SMS as an examination of dozens of messaging technologies *'with a large user base – and thus a great deal of sensitive user communication – in addition to smaller companies that are pioneering advanced security practices'*, implicitly involving both established and emerging actors, with arguably very different levels of security and usability, in the effort. The visual appearance of the SMS is presented in **Figure 1**: a simple table listing specific tools vertically and the seven classification criteria horizontally. These are the following:[11]

1. *Is your communication encrypted in transit?*
2. *Is your communication encrypted with a key the provider doesn't have access to?*
3. *Can you independently verify your correspondent's identity?*
4. *Are past communications secure if your keys are stolen?*
5. *Is the code open to independent review?*
6. *Is the crypto design well documented?*
7. *Has there been an independent security audit?*

The table includes intuitive symbology and colours to account for the presence or the lack of a requirement, and a filter giving the possibility to display only specific tools, or get a bird's eye view of all the examined tools in alphabetical order.

[10] Ibid.

[11] And we will come back to them extensively.



Secure Messaging Scorecard

Filter

- Any -

	Encrypted in transit?	Encrypted so the provider can't read it?	Can you verify contacts' identities?	Are past comms secure if your keys are stolen?	Is the code open to independent review?	Is security design properly documented?	Has there been any recent code audit?
AIM	✓	⊘	⊘	⊘	⊘	⊘	⊘
BlackBerry Messenger	✓	⊘	⊘	⊘	⊘	⊘	⊘
BlackBerry Protected	✓	✓	✓	⊘	⊘	✓	✓
ChatSecure + Orbot	✓	✓	✓	✓	✓	✓	✓
Ebuddy XMS	✓	⊘	⊘	⊘	⊘	⊘	⊘
Facebook chat	✓	⊘	⊘	⊘	⊘	⊘	✓

Figure 1: The Secure Messaging Scorecard Version 1.0.

Thus, as we can see, the political and 'quasi-philosophical rationale' behind the SMS is clearly presented. It is interesting to acknowledge, in this context, that the 'Methodology' section in the presentation page is very matter-of-factual, presenting the final result of the category and criteria selection, but very little information is shared by EFF on the thought processes and negotiations that led to the selection. The page states that *'Here are the criteria we looked at in assessing the security of various communication tools'* and proceeds to list them and their definition for the purpose of the exercise.

Yet, as researchers exploring the social and communicational dimensions of encryption from an STS perspective, we hypothesized that selecting and singling out these categories had been anything but linear, and the backstage of these choices was important to explore – not merely to assess the effectiveness of the SMS, but also as a way to explore the particular definition of encryption and security the EFF was inherently promoting as it was pushing the project forward. Indeed, it seemed to us that by fostering the SMS project, and due to its central role as an actor in the preservation of civil liberties on the Internet, the EFF was not merely acknowledging and trying to accumulate information on a 'state of things' in the field of encryption and security but it was contributing to shaping the field, the chosen categories to define 'actual' or 'usable' security being a very important part of *co-producing* it.

3.2. An ongoing reflection on the 'making of' the criteria

While the SMS's main presentation page focused on the criteria in their final form, there were some indications of the EFF's ongoing reflections on the 'making of' such criteria, and their presence on the Internet. An early November 2014 post by Chief Computer Scientist

Peter Eckersley, 'What Makes a Good Security Audit?', acknowledged that the foundation had 'gotten a lot of questions about the auditing column in the Scorecard'.[12] Eckersley proceeded to explain that obtaining knowledge on how security software has been reviewed for 'structural design problems and is being continuously audited for bugs and vulnerabilities in the code' was deemed an essential requirement, but it was also recognized that the quality and effectiveness of audits themselves could vary greatly and have significant security implications – opening up discussions that eventually led to the inclusion of three separate categories accounting for the code review aspect (regular and external audits, publication of detailed design document, publication of source code).[13] Despite the fact that methodologies for code audit vary, Eckersley gives examples of fundamental vulnerabilities that a 'good code audit' may decrease, such as Heartbleed or Shellshock. The 'code audit' criteria, thus, refer to other systems of classification, such as the CVE (Common Vulnerabilities and Exposures), an international standard for Informational Security Vulnerability names.[14] Other 'Notes' attached to some of the criteria (e.g. 4 and 5) are indicative of backstage negotiations, as they mention 'compromises' on forward secrecy (encryption with ephemeral keys, routinely and permanently deleted after the communication, a 'hybrid' version of which is accepted 'for this phase of the campaign') and on the source code's openness to independent review ('only require[d] for the tool and not for the entire OS').[15]

In addition, musings on the SMS came from several actors in the encryption/security community at large, some of them disputing the lack of hierarchy among the criteria, the phrasing of some of them, and even the messaging tools' alphabetical ranking,[16] others conducting alternative reviews on tools included in the SMS and disagreeing on their quality – despite the fact that they had scored positive on the EFF grid –[17] and yet others concluding that the EFF was being 'pushed to rethink' the SMS.[18] Indeed, relying on these documentary sources, it seems that reflections within the EFF on the quality of the SMS categories and their possible alternatives and evolutions on one hand, and more or less constructive external criticisms of the same aspects on the other, took place in parallel and paved the way for a subsequent version of the tool – all the while shaping 'good' security and 'good' encryption and how users and developers can have a proper grasp of them.

In early August 2016, these reflections seemed to have reached a turning point as the EFF 'archived' its SMS main page, labelled it a 'Version 1.0', announced it would be back in the near future with a new, improved and 'more nuanced' guide to encryption tools, and made an explicit reference to the ongoing revisions that had characterized and still were characterizing their categorization work: 'Though all of those criteria are necessary for a tool to be secure, they can't guarantee it; security is hard, and some aspects of it are hard to measure'[19] (**Figure 2**). It also anticipated that the term 'scorecard' would be dropped in favour of 'guide'.

[12] Eckersley, supra note 2.

[13] Ibid.

[14] http://cve.mitre.org/cve/.

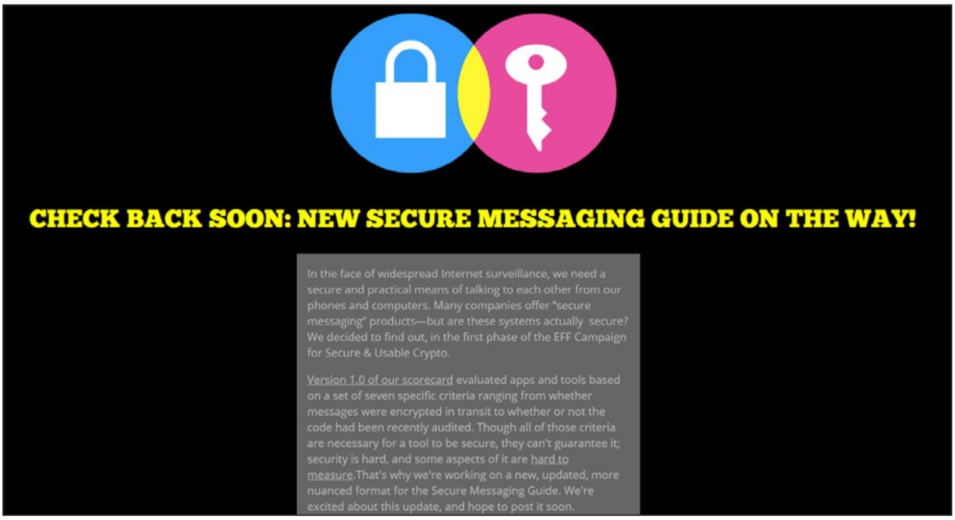
[15] https://www.eff.org/node/82654.

[16] Discussion thread on Hacker News, https://news.ycombinator.com/item?id=10526242.

[17] Daniel Hodson & Matt Jones, 'EFF secure messaging scorecard review', ELTTAM blog, 11 August 2016, https://www.elttam.com.au/blog/a-review-of-the-eff-secure-messaging-scorecard-pt2/.

[18] Zelijka Zorz, 'How the EFF was pushed to rethink its Secure Messaging Scorecard', HelpNetSecurity, August 11, 2016, https://www.helpnetsecurity.com/2016/08/11/eff-secure-messaging-scorecard/.

[19] https://www.eff.org/secure-messaging-scorecard.



CHECK BACK SOON: NEW SECURE MESSAGING GUIDE ON THE WAY!

In the face of widespread Internet surveillance, we need a secure and practical means of talking to each other from our phones and computers. Many companies offer "secure messaging" products—but are these systems actually secure? We decided to find out, in the first phase of the EFF Campaign for Secure & Usable Crypto.

Version 1.0 of our scorecard evaluated apps and tools based on a set of seven specific criteria ranging from whether messages were encrypted in transit to whether or not the code had been recently audited. Though all of those criteria are necessary for a tool to be secure, they can't guarantee it; security is hard, and some aspects of it are hard to measure.That's why we're working on a new, updated, more nuanced format for the Secure Messaging Guide. We're excited about this update, and hope to post it soon.

Figure 2: Secure messaging scorecard main page since August 2016 and as of February 8, 2017 (https://www.eff.org/secure-messaging-scorecard).

3.3. What did the first SMS do? Performativity and performance of a categorization tool

And thus, Version 1.0 of the SMS was no more, except in the form of an archive preserved 'for purely historical reasons' of which the EFF itself discourages further use.[20] Nonetheless, our interviews with EFF members and security trainers in several European countries – as well as the fact that the EFF decided to leave 1.0 visible to the public while putting it in context – demonstrate that the first SMS *did* a lot of things to and for the encryption and security community, and contributed to shaping the field itself.

Indeed, the first SMS appears to have had a performative effect on the community: the EFF has a central role as a protector of online civil liberties, and setting up a scorecard in this field was a pioneering effort in its own kind. In an otherwise mostly critical discussion thread among tech-savvy users,[21] it is recognized as such (*'The EFF scoreboard carries the embryo idea of a global crypto discussion, review, comparison and knowledge site that could also serve as a great resource for non-crypto people and students to learn a lot about that field'*), and most interestingly, it leads the encryption technical community to be reflexive about itself and its own practices. Firstly, because it was a 'hybrid' organization such as the EFF, which includes some technical people but a number of other profiles as well, that spearheaded this effort (*'Isn't it a bit strange that a small organization of non computer scientists produce something that was painfully missing for at least 50 years?'*), and secondly, because it prompts reflection on parallel categorization efforts that may better respond to what the technical community sees as 'good security' (*'The highly valued information you and other experts are dropping [...] should be visible in a place that collects all that stuff and allows for open discussion of these things in the public, so people can learn to decide what security means for them. If such a thing exists, please show me. If not, please build it.'*). Indeed, by its creation, the SMS has caused a

[20] https://www.eff.org/node/82654: *'you should not use this scorecard to evaluate the security of any of the listed tools'*.

[21] https://news.ycombinator.com/item?id=10526242. Citations in this paragraph are from this thread unless otherwise noted.

reaction in the developer community, making them ponder whether a categorization effort of this kind is worthy, and whether the EFF's particular categorization is. Despite the flaws they see in the scorecard, developers seem to perceive EFF as a sort of standardizing or trend-setting body. They know that many users will rely on the SMS if they perceive it to be the advice of EFF, but this is also due to its other initiatives aimed at bridging technical soundness and user-friendliness, such as the 'crypto usability' prize.[22]

Interestingly, and despite the EFF's announced intentions, we can retrace some early ambiguity and perhaps confusion amongst the SMS's intended target audience. The simplicity and linearity of the grid, the symbols used, the way categories were presented – all these aspects could indeed lead the encryption technical community to think that it was aimed at both developers and users (confirmed earlier by R1, as well), and perhaps primarily at users. However, there are other indications that it might have been the other way around – the primary target being to foster good security practices among a wide number of developer teams, and usability being an eventual target. According to R2, *'what motivated us to make the scorecard, is to survey the landscape of secure messaging and to show developers: 'look that's what we think is important. So you should use all these things to make a truly secure tool', and then we can get to the usability part'*. And later, even more clearly: *'originally the target of the SMS was not users, telling users 'you should use Signal or you should use something else'. [...] It was... we were trying to make a grid so that developers could see ok I need to get this and this to check all the boxes, but it backfired...'*. So, it appears that one of the core things the SMS did was to incite some practices by end users that escaped, to some extent, the EFF team's intentions and control: intended as an indicative checklist for developers, the SMS assumed the shape of an instruction tool for users and cryptography trainers – a 'stabilized' artefact, when in fact it was anything but, as the remainder of this section shows.

Ultimately, and despite the EFF's warnings,[23] developers appear worried that the SMS and its set of criteria will appear to guidance-seeking users as a performance standard that tools should achieve to qualify as 'good encryption' (e.g. tptacek on Hacker News: *'Software security is a new field, cryptographic software security is an even newer field, and mainstream cryptographic messaging software is newer still. The problem with this flawed list is that it in effect makes endorsements. It's better to have no criteria at all than a set that makes dangerously broken endorsements'*).[24] This concern is echoed by EFF, in R2's words during our interview: *'we were worried it was putting particular users whose safety depends on... you know users in Iraq or in Syria, in Egypt, where their security of their messages actually affects their safety. We were worried it was actually putting them in danger because they were interpreting one way while we had meant it another way'*. Indeed, the effect on 'real-life' uses of how SMS 1.0 was engineered and presented appears as one of the primary motivations to move towards a second version, something we will explore further in the last section of this article.

4. 'Security is hard to measure': Revisiting the SMS, (re-)defining security

> *We got a lot of feedback from security researchers who thought that it was far too simplified and that it was making some tools look good because they have hit all the checkmarks even though they were not actually good tools. So we ended up in a little bit in between*

[22] https://www.eff.org/deeplinks/2014/08/recap-first-eff-cup-workshop.

[23] Cf. https://www.eff.org/node/82654: 'the results in the scorecard below should not be read as endorsements of individual tools or guarantees of their security; they are merely indications that the projects are on the right track'.

[24] https://news.ycombinator.com/item?id=10526242.

> *zone, where it was not really simple enough for end users to really understand it correctly, but it was also too simple from an engineering standpoint.* – R1

In the interviews we conducted with them, EFF members describe how, since the early days of the SMS's first version, there has been an ongoing process of thinking back on its different categories, considering how they could be revisited, as well as the scorecard/'grid' itself as a device. Taking into consideration what actors in the encryption community considered 'errors' (shortcomings, misleading presentations, approximations, problematic inferences), and revisiting its own doubts and selection processes during the making of 1.0, the EFF team is currently analysing how the choice of these categories built towards specific understandings of encryption and security, considering how they could evolve – and with them, the definitions of 'good encryption' and 'good security' these categories present/suggest to the world. This third and last section addresses this ongoing process.

4.1. Questioning the grid: Incommensurability of criteria and the 'empty tier'

A first, fundamental level at which the reflection is taking place is the choice of the format itself. As this section's opening quote by R1 shows, the feedback on the first version of the SMS has revealed that in the attempt to be of use to both developers and users, the scorecard might have ended up as problematic for both. But this time, says R2, *'We want to make sure [...] that we can put out something that we're confident about, that is correct and is not going to confuse potential users'*. Especially in light of the meaning users bestowed upon SMS 1.0, is a grid the most useful and effective way to go? Perhaps it is the very idea of providing criteria or categories that is not suitable in this regard; actually, the updated project likely will not take the form of a grid, as R2 explains, to avoid the impression of prescription or instruction to users that the 'cutting up' of such a complex question in neat categories may previously have given:

> We are very likely to do the 2nd version of the scorecard but it is not a 100% sure thing at this point. And that's connected to the fact that we are completely changing the way the scorecard is set up, we are definitely abandoning the sort of grid of specific check boxes [...] A table seems to present cold hard facts in this very specific way. It is very easy to be convinced and it's very official [...] we are definitely going towards something more organic in that way, something that can capture a lot more nuance. [...] there's a lot more that makes a good tool besides six checkmarks.

Through the words of R1 and R2, we see how this intended additional 'nuance' is actually taking place not by eliminating categories, but by revisiting them. The v2 of the SMS, as currently envisaged, is supposed to divide the messaging tools in a set of different groups (R2 calls them 'tiers'): the first group will include tools that are unconditionally recommended, and the last group is meant to convey an 'avoid at all costs' message (e.g. for those tools that have no end-to-end encryption and present clear text on the wire). Within each group the tools would again be presented alphabetically, and not internally ranked; however, each of them is going to be accompanied by a paragraph describing it in more detail, instead of visual checkmarks, to the specific benefit of users: *'this is essentially so that we can differentiate... because we realize now that users are using this guide...'* says R2. And R1 suggests that this is also for the benefit of diverse user groups, with different levels of technical awareness: *'the goal is for people who are looking on the scheme on a very high level [to] say these tools are the best, these are bad, and there will be a slightly more nuanced explanation for people who really want to read the whole thing'*.

Interestingly, the EFF team plans to keep the first tier… empty. This has strong implications for the definition of good encryption and security – basically implying that it has not yet been achieved in the current reality of the secure messaging landscape, and as of yet, a mix of usability and strong security is still an ideal to struggle for: *'there's no tool we're aware of that is both easy to use and provides a high level insurance against state-level adversary'*. The new SMS will convey the message that nothing is actually 100 per cent secure, and make users aware of the fact that there is no 'perfect tool' yet to be recommended without reservation. R2 gives practical examples of why the team has come to this conclusion:

> There's nothing there because every app has some sort of a problem. WhatsApp has this issue of sharing data with Facebook so depending on your threat model is not great, Signal we thought a lot about reliability problems particularly outside the US, so it's not a great option. Those are scenarios where we'd say, those are maybe the best from what we can recommend, so they end up in 'adequate secure messaging tools' which is the next tier down, because they are the best tools that exist but we still want to emphasize to people that it's not 100% NSA-proof and you will never be safe for life, because we think that's overselling them.

In passing, the EFF team is defining what according to them is a 'perfect' secure messaging tool *'not just be end2end, be reliable and not share data and stuff like that, but it's also have to protect from the metadata analysis… that's an unsolved theoretical problem but we're gonna put it out there'.* The emptiness of the tier is meant to send a strong message to begin with, but is of course bound to evolve: *'if a tool got pretty close, and did not provide perfect protection against metadata analysis, we still might put it up in that tier and say look, these people have made a really strong effort […] But so far, it is going to be empty. Just to emphasize that there's still plenty of distance for tools to go, even the best ones'*. The EFF does not plan to skip its 'recommender' function, however, as we will see later in more detail, the focus is going to be placed on contexts of use and on the different 'threat models' of various user groups. Thus, weaknesses *'may not be shown in a check-box'*.

The graphic and spatial organization of the second version of SMS radically differs from 1.0's table. First of all, moving from a table to a list implies a different way of working with the information and undermines the idea of a direct, linear and quantified comparison offered by the table. A list of tiers with nuanced descriptions of different apps and their properties offers room for detailed and qualitative explanations, while tables tend to put different tools on the same surface, thus creating an illusion of commensurability and immediate quantified comparison. As R2 says:

> It's definitely gonna be a list-like, it's definitely not gonna be a table. We may or may not allow people to filter on some of these criteria. For example if you wanna see just iPhone ones we may put a little button where you can click on the top and only show only iPhone tools or only show tools that use Axolotl or something like that if you wanna to. But it will be more like a list: here is the first group, here's the next group, here's the next group, here's the group you should never use.

The idea of 'filters' adds a certain degree of user agency in the classification of tools: the lists become modulable as users may set up a criteria that could graphically reorganize data providing cross-tier comparisons. This offers a different way of both classifying the data and 'making sense of it', than a table. The latter is, as Jack Goody puts it, a graphically organized dataset with hard structure leaving little room for ambiguity (Goody, 1977).

Another problem posed by the grid format, that the new version of the SMS seeks to address, is the projected equivalence of the different criteria, from the open source release of the code, to its audit, to the type of encryption. The 'checklist' and 'points' system creates the impression of an artificial equality between these criteria, while in fact, the presence of some rather than others has different impacts on security and privacy, in particular for users in high-risk contexts:

> [If you were] someone who does not know anything about crypto or security, you would look at this scorecard and would say, any two tools that have, say, five out of seven checkmarks are probably about the same. When if… one of those tools actually had end 2 end encryption and the other did not, even though they both had code audit or the code was available, or things like that, their security obviously isn't the same. That was basically the flaw we found, the flaw that existed in using **that** system to present information to users because it artificially made the apps that were definitely not secure in the same way look secure. And we were worried it was putting particular users whose safety depends on… you know users in Iraq or in Syria, in Egypt, where their security of their messages actually affects their safety.

The EFF team also realizes that some developers and firms proposing secure messaging tools have, while presenting their tools, bent the grid to their own advantage – more precisely, they have presented a high conformity to the SMS as a label of legitimacy. Again, R2 emphasizes that this is a problem particularly in those cases when users lack the technical background or expertise to build their own hierarchy of the criteria's relative importance, according to their needs or threat model (see 4.2):

> [There] were tools that just said on their website that say we got 5 out of 6 on EFF SMS so we are great […] some of the tools tried to sneak advantage out of it, and I can't really blame them because it was designed in this way. […] For many users […] Even if you know what end2end means but you don't necessarily know… you're not in this realm… is having a code audit more important than, say, forward secrecy? […] the answer of course is that it depends on your "threat model". But there's nothing about the threat model or what you're protecting against in the [1.0] scorecard, which is part of the problem.

A piece of recent research on user understanding of the SMS, undertaken by Ruba Abu-Salma, Joe Bonneau and other researchers with the collaboration of University College London, shows that indeed, users seem to have misunderstood the scorecard in several respects. Four out of seven criteria raise issues: 'participants did not appreciate the difference between point-to-point and E2E encryption, and did not comprehend forward secrecy or fingerprint verification' (Abu-Salma et al., 2017), while the other three properties (documentation, open-source code and security audits) 'were considered to be negative security properties, with users believing security requires obscurity'. Thus, the assessment concludes, 'there is a gap not only between users' understanding of secure communication tools and the technical reality, but also a gap between real users and how the security research community imagines them' (ibid).

4.2. 'Not everyone needs a bunker!' From a tool-centred to a context-centred approach

The additional expertise needed by the user to understand the difference between various criteria and their importance emerges as a crucial flaw of the v1 grid. One of the keys for a 'new and improved' SMS, thus, seems to be the fact of taking user knowledge seriously, as a cornerstone of the categorization: for it to be meaningful, users must identify and analyse their respective 'threat model' – i.e. identify, enumerate and prioritize potential threats in their environment, in this case digital environment. R3 remarks that this is one of the core objectives the EFF in several of its projects beyond the revision of the SMS – and notes that there are no direct tools to indicate which threat model one has, but users need to uncover the right indicators in their specific contexts of action:

> We're not answering what someone's threat model is, we just help guide them in their direction of what to read. We can say like journalists might have good secure communication tools that they might wanna protect their data, but we can't say how much of a threat any journalists are under because different journalists have different threats.

R2 points out that the same words, used to identify a particular threat, might have very different meanings or implications depending on the geopolitical context – the more 'qualitative', descriptive part of the new tier-organized SMS should be useful to trigger the right reflexes in this regard:

> It's harder to say in 3 words what is your threat model. [...] even if it's easy to distinguish a threat model "I just want more privacy" versus "I am worried of a state-level actor", there's still a difference between "I am worried of a state-level actor in Syria" vs. "I am worried of a state-level actor in Iran" versus China versus US… Am I worried about all of them or a particular actor? I think it's a lot harder to capture it in the grid. We're hoping that tiers will capture that to some degree.

There is no universally appropriate application, so the new SMS should take the diversity of the users – and the corresponding diversity of their threat models – as a starting point. R2 again resorts to specific case-examples to illustrate this point:

> There's also totally different types of people who are coming to this thing. Some of them are Ukrainian journalists working in a war zone, and some people are hipster San-Francisco wealthy middle-class people in the US who don't really have to worry about the safety of their lives being dependent on their messages, but still wanna do something, still value privacy, and we might recommend Signal for one and WhatsApp for the others because they're already on Facebook so it does not really change anything for them, Facebook already knows everything about them, so it already has their contacts. So including this information and being able to tell people look you don't really have to completely change your life if your threat model isn't demanding it, you can use this tool or here's the slight change you can do. I think it's helpful, it makes more sense. Not everyone has to put a tin foil hat and create an emergency bunker. Lots of people do, but not everybody. Tailoring it to the right people. I think that would be great to have an app that we would recommend to everyone because it's usable and easy, popular and secure and everything else but since it's not there I think it's useful to tailor things, tailor the threat model.

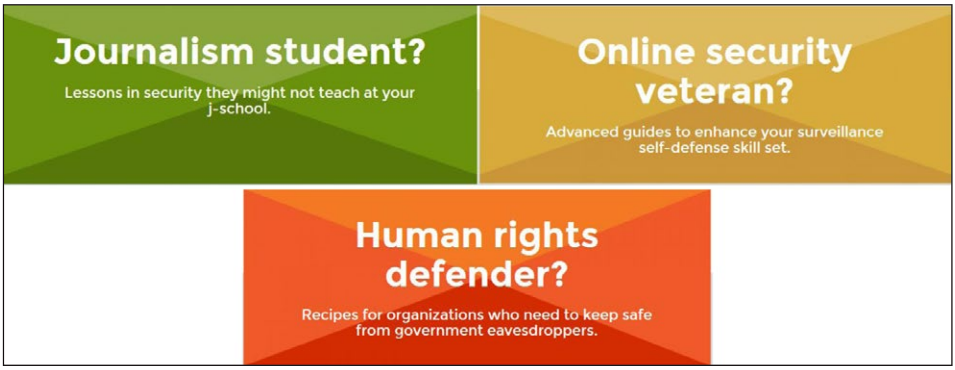
If, for the user, the choice of a strong secure messaging tool is in this vision strictly linked to the understanding of his or her threat-model, for its part the EFF acknowledges that if there is no universally appropriate application, so it goes for the definition of what constitutes 'good' encryption – 'good' security and privacy. Beyond the strength of specific technical components, the qualities of being 'secure' and 'private' extend to an appreciation of the geographical and political situation, of the user's knowledge and expertise of whether privacy and security are or are not a matter of physical and emotional integrity, which can only be contextually defined and linked to a particular threat model. A high-quality secure messaging tool may not necessarily always be the one that provides the strongest privacy, but the one that empowers users to achieve precisely the level of privacy they need.

4.3. Comparing with other categorization systems

The efforts to revise the SMS cannot, for the EFF team, do without a comparison with other categorization systems. On the one hand, the new version of the SMS will interact with the Surveillance Self-Defense Guide, developed by the EFF itself and designed to aid in 'defending yourself and your friends from surveillance by using secure technology and developing careful practices.'[25] Indeed, the contextual approach to users' needs and threat models appears to be dominant in this project: in stark contrast to the technical properties-based criteria of the first SMS, on the guide's home page, a number of 'buttons' introduce the reader to different paths of privacy and security protection depending on user profiles and 'idealtypes' (**Figures 3, 4** and **5**). The revised SMS should partake in this shift.

According to R3, this approach based on facilitation, induction and personalization informs more broadly the recent EFF efforts, and goes back to identifying the right level of relative security for the right context:

> 'We don't distinguish threat models, we give tools to help users figure out what are their threat models. [...] I still would not say we were putting an answer to the question out there. The key to the guide that we've created is that we want people to start with **understanding their own personal situation,** so their threat-model, rather than saying them just use these tools, I don't think that's a productive guide. [...] WhatsApp for example, it has end to end encryption. It may be good for an average person to just keep using that if they are already using it and learn how to use it well and correctly. But I think other people have much more extreme threat-models and have to use more secure tools'.



Journalism student?

Lessons in security they might not teach at your j-school.

Online security veteran?

Advanced guides to enhance your surveillance self-defense skill set.

Human rights defender?

Recipes for organizations who need to keep safe from government eavesdroppers.

Figures 3, 4 and 5: Sample of 'user paths' on the SSD home page (https://ssd.eff.org/en).

[25] https://ssd.eff.org/en.

The EFF is also looking at categorization systems in the same field produced by other actors – acknowledging that SMS 1.0 has been, in turn, an inspiration for some of them and that there seems to be a need, generally identified in the field, of tools providing guidance in the increasingly complex landscape of secure messaging.[26] According to R1, *'I talked to a lot of people that did projects like that, and a lot of them have taken some info from the old scorecard and some concepts from it [...] it's good to have different people taking a different take on and I think that would be bad there was only one attempt to do this'*. Both R1 and R2 refer in particular to the similar, recent effort by Amnesty International,[27] and while R1 acknowledges that *'they were really trying to produce something very simple and consumer-based and I think they did that, their report was much easier to digest for the general public'*, R2 mentions how *'to some degree I feel they suffer a lot from the same problem that Scorecard had, which is they try to produce things that rely to a single number, they gave this score like 90.5 points out of a 100 and that was exactly what we thought was the problem of the first scorecard: you can't reduce security down to a single number or a single set of checkboxes'*.

The alternative, user-centred approach arises not only as a result of internal reflections on the first version of SMS, but also because of how parallel categorization attempts were done by other actors promoting online civil liberties. Our recent interview with an informational security trainer also shows a turn in pedagogical methods from tools to threat model evaluation: *'Very often trainings turn into tool-trainings. But in our work tools are not our primary and even not secondary concerns. What's primary is the evaluation of what participants need, what they already use. And only after we think of what we can suggest them to use, and again, without any hard recommendations 'you need only this tool and that's all'*. [Informational security trainer, Ukraine].

4.4. Conformity to criteria: Evidence-seeking, evidence-giving

A final set of evolutions moves from the building of the categorization system itself to how proof of the tools' conformity to the guidance provided will be requested by EFF and provided by the developers of the tools, and the encryption/security community at large. Indeed, part of the early criticisms of the first SMS did not have to do with the format of the grid, but were due to the opacity of the ways in which its 'binary' recommendations were evidence-supported.

In SMS 1.0, 'green lights' had sometimes been awarded for specific criteria as a result of the private correspondence between R1 and the developers, says R2: *'He would just email [...] sometimes he knew who was the developer but most cases it was just like contact @...'*. However, the new version is going to adopt a more transparent approach and encourage display of public evidence from the developers, the lack of which may be a deal-breaker: *'this time around we are not going to accept as proof of any criteria any private correspondence. If an app wants to get credit for something, it has to publically post it somewhere. I mean we may be contacting developers to encourage them to publically post it. [...but] we don't want to have to say, well, we talked to them. We want to say they are publicly committed to it'*.

In particular, one of the criteria that had raised more objections in terms of provided evidence was the review of the code ('What does "security design properly documented" even

[26] This was also brought up in the Hacker News SMS-related thread: 'Isn't it a bit strange, that there is no such thing as that scoreboard produced by an international group of universities and industry experts, with a transparent documentation of the review process and plenty of room for discussion of different paradigms?' (see https://news.ycombinator.com/item?id=10526242).

[27] https://www.amnesty.org/en/latest/campaigns/2016/10/which-messaging-apps-best-protect-your-privacy/.

mean?', a developer had commented).[28] The EFF did not have sufficient resources to review the code line by line, and for several of the tools the code was not available, which is why the 'external audit' criterion was added. The lack of resources to dedicate to this task is still a problem, one more reason, according to R2, *'why the second phase is basically looking at what does the app developer publicly say or put on the website'*. In parallel, the criterion calling for an independent audit – which, as we recall, had elicited a lot of internal methodological reflection since the beginning – will no longer have a place in the second version of the SMS, once again because of evidence-seeking requirements. As R1 points out, *'the cost of really doing a crypto audit of a tool is really high. The audit of 40 tools would have taken an entire year for me. So we just did not have the resources'.* Interestingly – and while he agrees that the criterion needs to be dropped as such – R2 points out that several companies do their own audits, and the fact that they keep the results private does not necessarily affect their quality: *'there are companies like Apple or Facebook [...] it's almost certain they're doing an audit with an internal team so they will not release it publically. It does not necessarily mean that internal team did not do a good job. [...] For that reason we feel like the whole audit category does not do a lot. But we are going to still include if the code is open for an independent review because we think that's important [for some threat models]'*.

Finally, to support the arguments provided in the new SMS, the EFF team hopes to build on the competencies of the encryption community of practice – cryptographers, professors, people in industry and trainers. R2 remarks: *'We hope we're going to do this more for the 2nd version of scorecard, get feedback from cryptographers and people who are experts in this area to find out... did we miss something, are we categorizing the app wrong because we missed something'.* The search for feedback will be ongoing, to avoid falling into the same trap that had led commentators to wonder why the making of the first SMS had seemingly been a mostly 'internal' matter for inset 'the' EFF,[29] and also to be able to promptly react to important changes in the tools: *'We try to make it clear that we're keeping the door open to feedback after we publish it. For anyone we did not get a chance to talk to before we published it. We can't get feedback from everybody before it goes live'.* As tools evolve, what constitutes 'good' security evolves, or may evolve, as well.

5. Conclusions

This article has sought to examine the role of categorization and classification attempts in both the first and the foreseen second version of the SMS. In doing so, it has analysed how, by challenging, re-examining and re-shaping the categories that are meaningful to define the quality of secure messaging tools, the EFF has sparked a 'global crypto discussion'[30] that currently contributes to shape what constitutes 'good' security and privacy in the field of encrypted messaging.

Indeed, on one hand, the EFF's activities, epitomized by the SMS and its revisions, seem to contribute to the 'opportunistic turn' in encryption (IETF, 2014) that gained momentum in 2014 after the Snowden revelations, and consists of a progressive move of the crypto community towards making encryption 'seamless', with almost no efforts required from users. In terms of design choices, this entails a 'blackboxing' of quite a few operations that used to be visible to users, and needed to be actively controlled by users (e.g. key exchange and

[28] Ibid.

[29] See https://news.ycombinator.com/item?id=10526242: 'Why didn't they consult any named outside experts? They could have gotten the help if they needed it; instead, they developed this program in secret and launched it all at once'.

[30] Ibid.

verification, choice of encrypted/unencrypted status etc.). The opportunistic turn calls for an 'encryption by design', and constructs a new user profile, one who 'does not have to' have any specific knowledge about cryptographic concepts and does not have to undertake any additional operations to guarantee a secure communication. That shift may also be explained by the growing popularity of 'usable crypto' that undermines experts' monopoly and makes easy end-to-end encryption accessible outside of the tech-savvy user groups, where users used to be at the same time designers, developers and cryptographers.

However, while calling upon developers for improved usability – demanding that the technical crypto community make some properties, such as key verification, easy for users and independent from user agency – the EFF puts users at the core of the new categorization system, and in doing so, entrusts them with an important decision-making responsibility. To put it in R2's words, *'[The aim] is still to push the developers to improve but we realize that people are using it to make choices, so now the idea is [that] instead of just showing the developers here's what you have to do'*; as it is often the case with categorization and classification systems, the SMS was re-appropriated by the different actors in the community of practice, beyond the intentions of its creators – and in particular, several users have relied on it heavily. The new version intends to take this into account to a larger extent; but within the new paradigm built into the design of the second version of SMS, users have to question their threat models, increase their awareness of them, and have to know how to make technological choices according to their particular situation – a tool is 'good' if pertinent to the context of use. This paradigm shift is also experienced and reflected upon by trainers, organizers of cryptoparties and informational security seminars whom we have interviewed in different countries.

In the end, according to R1, a 'good' secure messaging tool *'would be something that security people would be fine using and also like people were just naturally using because it was a good product. [...] Security would not have to be something people thought about, and you would not have to be switched to a special app to have a secure conversation, but it will just be a default'*. It is this working definition of 'good' security and privacy, reminiscent of the privacy-by-design and privacy-by-default approaches (Cavoukian, 2012; Willis, 2014) that is gaining increasing relevance, that the EFF is constructing with the SMS negotiations and related efforts – one likely to be achieved by tools that merge technical soundness and usability.

Acknowledgements

This work is supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation (H2020-ICT-2015, ICT-10-2015) under grant agreement n° 688722 – NEXTLEAP.

Competing Interests

The authors have no competing interests to declare.

References

Abu-Salma, R., et al. (2017). Obstacles to the adoption of secure communication tools. In: *Proceedings of the 38th IEEE Symposium on Security and Privacy*. (Oakland), San Jose, CA, USA. DOI: https://doi.org/10.1109/SP.2017.65

Aouragh, M., et al. (2015). Let's first get things done! On division of labour and techno-political practices of delegation in times of crisis. *The Fibreculture Journal*, 26. DOI: https://doi.org/10.15307/fcj.26.196.2015

Becker, H. S. (1986). *Doing Things Together: Selected Papers*. Evanston, IL: Northwestern University Press.

Bowker, G., & **Star, S. L.** (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.

Callon, M. (2009). Elaborating the notion of performativity. *Le Libellio d'Aegis*, 5(1): 18–29.

Cavoukian, A. (2012). Privacy by design [leading edge]. *IEEE Technology and Society Magazine*, 31(4): 18–19. DOI: https://doi.org/10.1109/MTS.2012.2225459

Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Harvard University Press.

Denis, J. (2006). Les nouveaux visages de la performativité. *Études de communication*, 29: 8–24.

Ermoshina, K., Halpin, H., & **Musiani, F.** (2017). Can Johnny build a protocol? Co-ordinating developer and user intentions for privacy-enhanced secure messaging protocols. *Proceedings of the 2nd European Workshop on Usable Security*. Internet Society. Available at: https://www.internetsociety.org/sites/default/files/eurousec2017_16_Ermoshina_paper.pdf.

Ermoshina, K., Musiani, F., & **Halpin, H.** (2016). End-to-end encrypted messaging protocols: An overview. In: Bagnoli, F., et al. (eds.), *Proceedings of the Internet Science Third International Conference, INSCI 2016*, 244–54. Florence, Italy, 12–14 September, Springer. DOI: https://doi.org/10.1007/978-3-319-45982-0_22

Froomkin, D., & **McLaughlin, J.** (2016). FBI vs. Apple establishes a new phase of the crypto wars. *The Intercept*, 26 February. Available at: https://theintercept.com/2016/02/26/fbi-vs-apple-post-crypto-wars.

Goodwin, C. (1996). Practices of color classification. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 3(2): 62–82.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Gürses, S., Kundnani, A., & **Van Hoboken, J.** (2016). Crypto and empire: The contradictions of counter-surveillance advocacy. *Media, Culture and Society*, 38(4): 576–90. DOI: https://doi.org/10.1177/0163443716643006

Internet Engineering Task Force. (2014). Request for Comments 7435, Opportunistic Security: Some Protection Most of the Time. Available at: https://tools.ietf.org/html/rfc7435.

Kazansky, B. (2015). Privacy, responsibility, and human rights activism. *The Fibreculture Journal*, 26. DOI: https://doi.org/10.15307/fcj.26.195.2015

Lave, J., & **Wenger, E.** (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511815355

Musiani, F., et al. (eds.) (2016). *The Turn to Infrastructure in Internet Governance*. New York: Palgrave Macmillan. DOI: https://doi.org/10.1057/9781137483591

Rubinstein, I., & **van Hoboken, J.** (2014). Privacy and security in the cloud: Some realism about technical solutions to transnational surveillance in the post-Snowden era. *NYU School of Law, Public Law Research Paper No. 14–46*. Available at: https://ssrn.com/abstract=2443604.

Sanger, D., & **Perlroth, N.** (2015). Encrypted messaging apps face new scrutiny over possible role in Paris attacks. *New York Times*. Available at: http://www.nytimes.com/2015/11/17/world/europe/encrypted-messaging-apps-face-new-scrutiny-over-possible-role-in-paris-attacks.html.

Unger, N., et al. (2015). SoK: Secure messaging. In: *2015 IEEE Symposium on Security and Privacy*, 232–49. IEEE. DOI: https://doi.org/10.1109/SP.2015.22

Willis, L. E. (2014). Why not privacy by default? *Berkeley Technology and Law Journal*, 29: 61–133.

How to cite this article: Musiani, F. and Ermoshina, K. (2017). What is a *Good* Secure Messaging Tool? The EFF Secure Messaging Scorecard and the Shaping of Digital (Usable) Security. *Westminster Papers in Communication and Culture,* 12(3), 51–71, DOI: https://doi.org/10.16997/wpcc.265

Submitted: 11 February 2017 **Accepted:** 26 September 2017 **Published:** 31 October 2017

]u[*Westminster Papers in Communication and Culture* is a peer-reviewed open access journal published by University of Westminster Press

OPEN ACCESS