The Authentication Equation: A Tool to Visualize the Convergence of Security and Usability of Text-Based Passwords

Cathryn A. Ploehn(✉) and Kristen K. Greene

National Institute of Standards and Technology, 100 Bureau Dr, Gaithersburg, MD, USA
{cathryn.ploehn,kristen.greene}@nist.gov

Abstract. Password management is a ubiquitous struggle of the modern human. Despite usability playing a vital role in authentication, many password policies and requirements focus on security without sufficient consideration of human factors. In fact, security and usability needs are often in contention. Until an improved authentication method beyond character input is implemented on a large scale, developing new methodologies for balancing competing requirements is vital.

This research project focused on building a data visualization tool to explore password usability and security metrics. The visualization tool integrates various measurements of passwords, enabling exploration of the intersection of their usability and security components. The tool is based on insight from previously gathered data from usability studies conducted at the United States National Institute of Standards and Technology. It also leverages web technologies to flexibly display data sets computed from sets of passwords. The tool is available at https://github.com/usnistgov/DataVis.

Keywords: Data visualization · Usable security · Keystrokes · Entropy · Password policies · Password permutation

1 Introduction

There is an abundance of usability failures of text-based passwords. Passwords add to the mental strain of users and are managed by counterintuitive requirements. Both personal and work password-protected accounts are impacted by such requirements. Learning and recalling complex passwords for multiple accounts uses mental resources and time [1–4] that can be better applied elsewhere.

Our current research is focused on password requirements from the enterprise perspective. As a result of the cognitive load password management requires, employees must learn to cope with unusable passwords [4–6]. Recent research has surveyed the coping strategies of different groups, such as university members (students and staff), federal employees, and employees in other types of enterprises with regards to password management [1,5–7]. Potentially compromising

T. Tryfonas and I. Askoxylakis (Eds.): HAS 2015, LNCS 9190, pp. 95–106, 2015.
DOI: 10.1007/978-3-319-20376-8_9

password management strategies employees engage in include: using a previous password with minor changes, using an existing password, recycling an old password, using a common name, and using at least one storing method [1,6–8]. Some employees have a false sense of security around their work-related accounts and passwords [1,5,9]. In shifting the locus of control away from their actions and towards the perceived security of the system, employees may continue to use insecure password generation and maintenance methods without the necessary self-scrutiny.

The password policies enforced in many institutions exacerbate the usability weaknesses of passwords, eliciting negative attitudes from employees [7,8]. A recent survey of United States Department of Commerce employees indicates a correlation between employee's negative attitudes towards text-based authentication and the competency of their resulting password management behaviors [1]. Furthermore, employees surveyed felt the passwords required were too long (56.9 %) and too complex (50.7 %) [1]. Previous research reinforces the idea that long, complex character strings are harder to recall and more error prone due to the increased cognitive load [2,9]. The difficulty with using complex character strings is exacerbated on mobile devices due to the constraints of smaller onscreen keyboards [10].

It is widely accepted that the use of text-based passwords is not the ideal authentication mechanism. Research is being done to re-imagine the methods we use to authenticate with the ideal balance of security and usability [11,12]. However, an improved authentication paradigm, such as that envisioned by the National Strategy for Trusted Identities in Cyberspace [13], will take some time to become implemented on a large scale.

In order to improve the usability of text-based passwords in the shorter term, the specific pitfalls of passwords should be explored. An identification of which specific aspects of text-based passwords impact usability is vital. Only with a solid understanding of the mechanisms that affect the usability of passwords can their management be improved. Password requirements can be improved based on this understanding to alleviate some of the cognitive load on password users.

Interactive visual analysis can be an invaluable tool in the pursuit of a more usable password. A quality visualization paradigm can aid in unearthing hidden relationships within a data set, supporting the analysis of large quantities of data very quickly [14]. Ben Schneiderman's **information seeking mantra** is a useful guideline for gleaning insight from data [15]: *Overview first, zoom and filter, then details on demand*. We utilize this concept to guide the design of our own interactive visualization tool.

The main research goal of this project was to facilitate NIST's[1] exploration of where the security and usability of passwords intersect. Although intended to specifically support NIST research, the tool is also available to the wider research community. Reaching this goal involved two major activities: (1) identifying measurable password components, or metrics, to analyze and automating the computation of these metrics for sets of passwords and (2) building a

[1] National Institute of Standards and Technology.

visualization tool to allow the dynamic exploration of the computed data sets. The tool should facilitate the comparison of usability and security metrics for sets of system generated passwords[2] and allow dynamic exploration of many different data sets. The visualization tool should also aid in the determination of which password components are significant in regards to usability and security, driving the collection of new data or the formulation of new password metrics for further study. The tool should ultimately give insight into how to better manage passwords with regards to organizational password requirements and password generation.

2 Methodology

2.1 Identification of Password Metrics

The password components initially selected for representation in the tool centered around common measures of security and previous NIST work on usability of system generated passwords. Metrics in the current iteration of the automation code include: linguistic and phonological difficulty (LPD), number of keystrokes, and entropy.

Linguistic and Phonological Difficulty Score. In previous research, a linguistic and phonological difficulty (LPD) scoring system was developed to rate the usability of passwords based on their similarity to spoken or written language patterns [16]. A difficulty score is generated based on the scores of six sub steps: whether the password begins with a symbol; the number of chunks (groups of numbers or letters separated by symbols) a password has; the size of the chunks; whether any letters are capitalized within a chunk; whether the letters, numbers, and symbols are segregated in each chunk or mixed together; and whether the password is pronounceable [16].

Keystrokes. The number of keystrokes needed to enter each password is also measured. For demonstration purposes, the landscape keyboard of the Android Galaxy 3s[3] and the landscape keyboard of the iPad 3 were used to calculate mobile keystrokes. The proliferation of non-desktop devices requiring password entry adds another layer of potential error to the usability equation [17]. Furthermore, due to the introduction of a touchscreen keyboard, symbols are buried in

[2] In contrast to the variability that exists in human generated passwords, system generated passwords can be created with more control. Multiple sets of system generated passwords were readily available from previous research. Therefore, system generated passwords were used as a starting point in the current work with a future goal of investigating user generated passwords.

[3] Disclaimer: Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

multiple screens instead of persistently visible as they are on desktop keyboards. Depending on the form factor, users need to navigate through multiple screens to type a single character, adding extra keystrokes depending on the device type and operating system. Multiple screens and additional keystrokes add a layer of cognitive overhead to the authentication process, rendering device type a definite usability factor for text-based authentication [10].

Entropy. Metrics for measuring password security fall into two main groups based on how a password was created: user generated or system generated. As the tool was initially created for use with system generated passwords, entropy is the measure of security used in the current iteration of the tool. Information entropy, or randomness, is commonly used to measure password strength for system generated passwords. Use of the term entropy in information theory was coined by Claude Shannon [18]. As bits of entropy measured in a password increase, the predicted measure of security increases. For the purposes of the current iteration of this tool, we used a general formula for entropy from NIST Special Publication 800-63-2, Appendix A [19].

Password Permutation. Password permutation has been suggested as a means of improving complex password entry on mobile devices [20]. By rearranging, or permuting, passwords such that like character classes (i.e., uppercase, lowercase, numbers, and symbols) are grouped together within a password, it reduces the number of keystrokes required to enter the password. Keystrokes are reduced since the user does not have to continually switch back and forth between multiple onscreen keyboards to access the numbers and symbols.

For the sake of the visualization, all previously mentioned metrics were computed for both the original passwords and their permuted counterparts, including: LPD, entropy, and number of keystrokes. It is important to note that the method of computing entropy had to be revised for the permuted passwords, as restructuring the passwords in a predictable format of uppercase, lowercase, numbers, and symbols reduces entropy [20].

2.2 Challenges Automating Metric Computation

The first component of the visualization tool is code that computes usability and security metrics from lists of passwords. The code is written in Python v3.4.0, using a text file as input and a comma separated values file as output.

In the design of the code to compute password metrics, many ambiguities and design questions arose. It was necessary to translate the previously designed LPD score from natural language into a consistent formal language equivalent in order to automate the computation of this score.

Calculating entropy also presented an interesting challenge in terms of pre-permutation and post-permutation entropy, since rearrangement of the characters diminishes the resulting information entropy of the password [20].

2.3 Designing the Tool

As the literature evolves concerning password usability, a tool that is dynamic, customizable, and flexible is vital to unearth important relationships between password metrics.

The tool was designed with browser technology (HTML5, CSS3, and JavaScript) to maximize the flexibility, interactivity, and ease of dissemination of the tool[4]. The tool has a low barrier to entry since it is used on a platform independent browser. It can be utilized from any desktop machine on the Chrome browser. We leveraged D3.js, a JavaScript library enabling the altering of documents based on data assigned to different elements within the Document Object Model (which defines the structure of a document) [21].

Based on data sets already in NIST's possession, we determined the visualization tool required the following attributes: scalability for differing data set sizes, display of different tiers of granularity, comparison of different password metrics side by side, and the ability to interact with and customize the view of the data.

Scalability. The sets of NIST passwords to be visualized with the tool could be a range of sizes, in some cases exceeding thousands of passwords. Thus, the tool should accommodate for and display different sizes of data sets. According to Tufte's Shrink Principle, data graphics can (and often should) be shrunk down in size, increasing their data density [22]. To allow the differentiation of entire data sets with individual data points at very small sizes, a heatmap paradigm is employed. The heatmap is a familiar visualization methodology, allowing for easy comparison at varying levels of scale. Furthermore, displaying the data as a matrix prevents data points from being obscured by other data points, such as in parallel coordinates[5]. The columns of the grid represent specific password metrics (entropy, number of keystrokes, etc.). Each row of the grid represents a password, with each block on that row representing a specific usability or security metric of that particular password. Each block is colored darker or lighter according to the value associated with that specific metric. Darker colors indicate a greater value (e.g. higher numbers of keystrokes) while lighter colors indicate lower values (e.g. lower numbers of keystrokes).

Tiers of Granularity. Effective interactive visualizations allow analysis of data at a macro and a micro level. Patterns can arise at any level of granularity in a data set. Three tiers of granularity are provided in the tool, each with a slightly different representation of the data. Figure 1 shows an overview of the tool (including all three tiers). The first tier shows a miniature view of the entire data set (zoomed out). The second tier shows a neighborhood view of the dataset, with about 20 to 50 rows of adjacent passwords. This tier is scrollable, by using

[4] The source code for the tool can be found at https://github.com/usnistgov/DataVis.

[5] Parallel coordinates are commonly used to visualize multivariate data. Coordinate axes are placed in parallel with associated data points connected by lines.

the first tier as a scrollbar. The third tier shows password metric results on an individual level, with each metric value of an individual password displayed in detail. All different views of the dataset are simultaneously visible, allowing interactions with the data to include three contexts for the data points explored.

Side by Side Comparison. The grid paradigm provides a simple solution to visually compare different password metrics side by side. Different passwords can be compared based on the changing color saturation of the heatmap. Furthermore, symmetry, particularly bifold reflective symmetry, has a perceptual immediacy in the human mind [23]. We designed the tool to capitalize on this fact, placing the metrics of the permuted passwords to the right of the original password metrics in reverse order (Fig. 1). The locations of the metrics are symmetrical to one another when comparing the original passwords with their permuted counterparts in the grid, allowing different levels of change to be easily detected based on their levels of symmetry.

Interactivity. The tool presents a non-static view of the data, enabling users to change the view of the data based on their interaction. When the mouse hovers over a particular block in the second tier of the visualization, the row and column of that selected block changes color scheme for ease of comparison. The context of any given data point is reinforced visually upon interaction. Furthermore, when blocks are hovered over in the second tier, the corresponding rows and columns are highlighted in the other tiers (Fig. 2). The need to explore patterns based on structural or other password characteristics necessitated the design requirement for dynamic sorting and filtering of passwords in the tool. The ability to subset and rearrange the view of large data sets is a powerful ability in order to hone in on patterns in the underlying data. We equipped the tool to dynamically rearrange, show, and hide grid columns and rows for a fully customizable view (Fig. 2). For example, the tool is equipped to filter passwords by length or by the amounts of numbers, letters, and/or symbols. Any sort and filter technique can be done in conjunction with other customizations on the fly, which update in real time with the tool. Filtering can be done using the controls within the accordion menus on the upper right hand side of the screen (Fig. 2). Using the powerful data selection and manipulation capabilities of D3.js, the amount of further customizations to the sort and filter capability of the tool is only limited by the number of calculations that can be done on the raw data [21].

3 Walkthough

We now give a brief walkthrough of browsing a dataset with the visualization tool using Fig. 2 as an example. The filter controls on the upper right hand side indicate the subset of passwords displayed. As indicated in Fig. 2, passwords of length 8 to 10 are displayed with all other password lengths filtered out. The passwords with the full range of letters are displayed (passwords containing 2 to

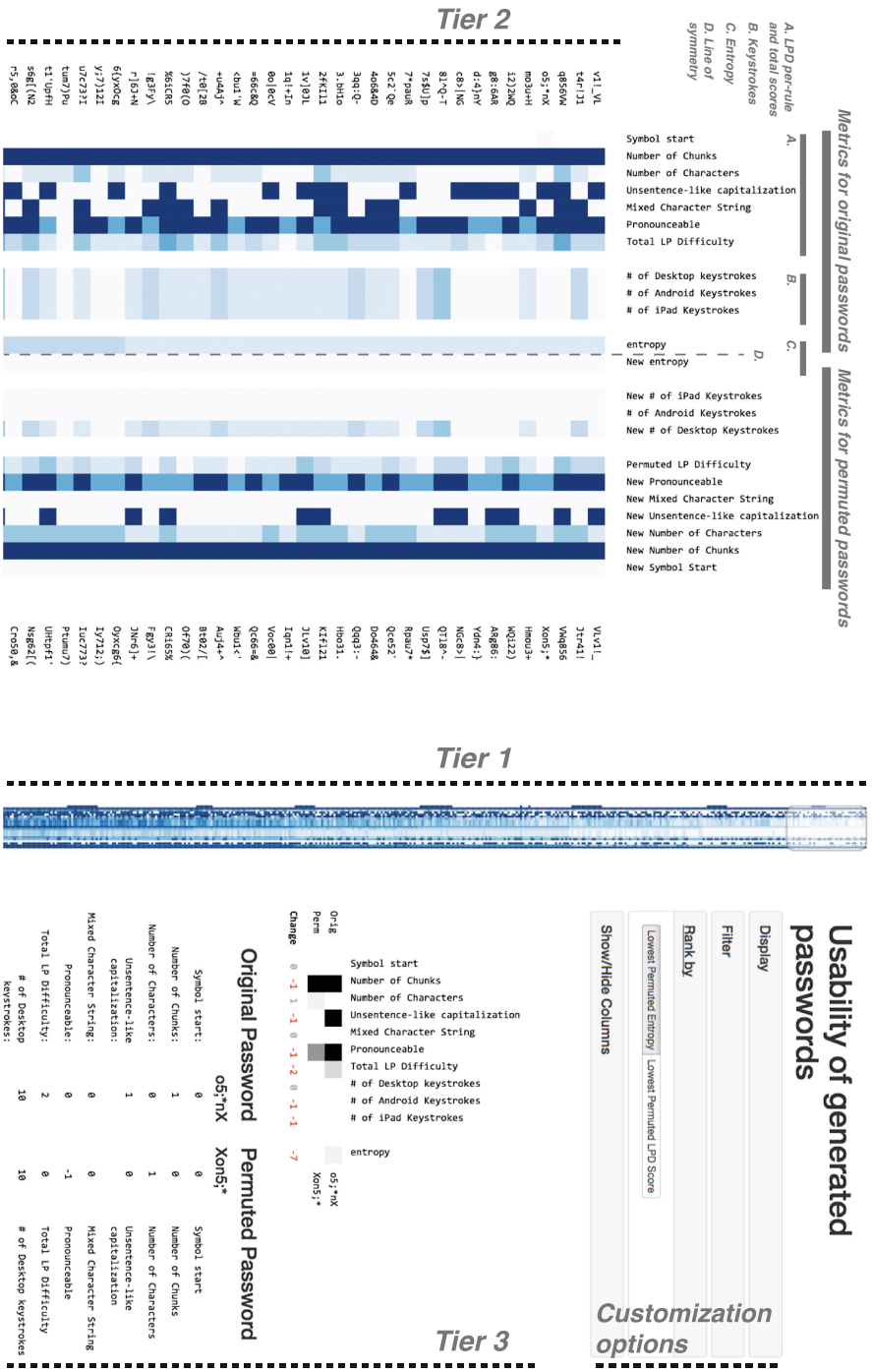

Fig. 1. The visualization tool (annotated) (Color figure online).

Fig. 2. Filtering with the tool (Color figure online).

10 letters). Passwords containing 1 to 4 numbers and 0 to 2 special characters are also displayed. The highlighted password is in the middle of this particular subset, which is ranked by lowest permuted entropy by default.

In Fig. 2, the metrics of the password `3f&{48D0C` and the metrics of its permuted counterpart, `DCf3480&{`, are highlighted across all three tiers as a result of a mouseover on the left hand side grid (the second tier). The first tier indicates the position of the specific password in the context of the filtered dataset at large. The second tier shows the context of the highlighted password at the neighborhood level. The third tier shows the specific metric scores for the password `3f&{48D0C` and its permuted counterpart, `DCf3480&{`. The rows and columns highlighted change as the mouse moves across the grid in the second tier. Therefore, in Fig. 2, the column "New Unsentence-like capitalization" is also highlighted.

On the bottom of the third tier, the number of keystrokes for Android and iPad decrease as expected as a result of the password permutation. Entropy also expectedly decreases. The interesting thing about this particular password is that the LPD score becomes higher after permutation. Looking at the third tier, the LPD score shows an increase from 3 to 4. In other words, a structural change that increases the mobile usability of this password (i.e. reduces the number of keystrokes necessary on mobile devices) also results in an increase to its LPD

score. Looking at the top of the third tier, the change from light to dark can be observed in "Number of Characters[6]" and "Unsentence-like Capitalization." The change in both totals up to an increase of 3 in the LPD. The "Number of Chunks" and "Mixed Character String" steps contribute to a decrease of 2 in the LPD at the same time, which brings the total gain of the permuted LPD score of this password to 1 (higher LPD scores mean greater linguistic and phonological difficulty, which should reduce usability).

A next step in analyzing the data could include a continuation of exploring whether the permutation of passwords in this dataset correlates with a decrease in LPD. Hiding all columns except for total LPD score would facilitate this exploration. This walkthrough demonstrates the ability to explore datasets across different levels of granularity.

4 Lessons Learned

During development, there were a few lessons learned regarding the usage of custom visualization tools to explore varying data sets. First, it is important to maintain a clear idea of how the data is manipulated and transformed into a visual display. Otherwise, a misinterpretation of the data is likely. For example, the color scale of the heatmap is generated on a case-by-case basis. The high and low scores for each metric of each data set visualized are used to map the values to colors. This means each metric has its own mapping for colors. Each data set has a custom color mapping, meaning dark colors and light colors are only high and low scores relative to the range of that particular data set. Two data sets cannot be directly visually compared with these color mappings. However, without knowledge of how the data is transformed into a visual representation, that discrepancy may be lost upon users of the tool. The same concept applies to similar visualizations for other experimental data sets. Without a thorough understanding of the mechanics of a tool, misguided conclusions can be gleaned from visualizations.

Second, special consideration is needed for visualization tools developed for a particular type of data source (e.g., system generated versus human generated passwords). Various data sets, although analogous in type, may not be fully compatible with different data tools. For example, our original tool computes entropy for system generated passwords. The concept of measuring information entropy in passwords is dependent on the randomness of the generation of these passwords. Subsequently, if passwords generated by humans were to be used in the same tool, the calculation of entropy would be inaccurate; measuring the entropy of human-generated passwords is a different calculation than measuring the entropy of system-generated passwords.

Underlying these lessons is a need to provide quantifiable evidence to support any assertion made based on a visualization. The goal of any visualization

[6] Note that the "Number of Characters" in the LPD rules refers to the number of characters within each chunk a password is divided into [16].

is to show the underlying data. Subsequently, any conclusions derived from a visualization must be backed up with the relevant data sources.

5 Future Work

5.1 Refine Metrics

We created the tool with the ability to evolve and change as the field's knowledge of usability and security regarding text-based authentication grows. As a next step NIST can use the tool to compare password usability metrics with human usability data (such as recall failures, input error rates, time to input, memorization times, etc.). For example, the LPD score could be compared with usability data collected from human subjects. If the human usability data did not correlate with the LPD score, the LPD metric could be revised or discarded from the tool altogether. Such data would inform the refinement of password metrics included in the tool for a more accurate representation of password usability.

As the field moves forward in defining useful password metrics to measure with regard to their impact on usability and security, the tool can easily be modified to accommodate and integrate these measures in its visual comparison paradigm. Entrenched in the tool's genesis is NIST's goal of ultimately understanding which usability components of passwords should inform password management policy.

Expansion of the security metrics included in the tool is another next step. Accurately describing password strength is vital. The trade-offs between theoretical methods and more realistic methods of security measurement have been discussed [24]. Other methods utilizing a source-independent measurement of password resistance to guessing have been developed [25–27], which could potentially be used for human generated and system generated data sets of passwords alike.

5.2 Additional Tool Functionality

The tool can be extended in a number of ways to facilitate and augment NIST's research in exploring password metrics. Analyzing data sets from multiple password generators or the human generated passwords resulting from different password policies side by side is currently impossible, for example, because the color scale used is generated based on the maximum and minimum data values within individual data sets. All data sets to be compared must be loaded into the tool for the color range to be generated correctly for direct comparison. The ability to explore and compare multiple data sets simultaneously would allow an investigator to better explore the trade-offs between one password generator and another, or between one set of password requirements and another.

Adding more visual paradigms to the tool would also serve a beneficial purpose. The grid paradigm is familiar, though the inclusion of other paradigms commonly used for multivariate analysis, such as parallel coordinates, would

add layers of sophistication to the level of visual analysis one can accomplish with the tool. It would be useful to see the underlying data displayed in different forms to unearth additional patterns that could be missed in one particular visualization method.

Streamlining the user experience would add a layer of needed usability to the tool. The current tool consists of two separate pieces of code in two languages, requiring a two-step process for users to visualize their passwords. Using the tool requires some knowledge of the command line and code editing, which could be encapsulated from the user and replaced with an easy to use, browser-based graphical user interface. Finally, the tool itself should undergo a more formal usability evaluation, having usable security researchers test it by exploring various data sets.

References

1. Choong, Y.Y., Theofanos, M., Liu, H.K.: United States Federal Employees Password Management Behaviors-a Department of Commerce Case Study. National Institute of Standards and Technology Interagency Report (NISTIR) (2014)
2. Stanton, B.C., Greene, K.K.: Character strings, memory and passwords: what a recall study can tell us. In: Tryfonas, T., Askoxylakis, I. (eds.) HAS 2014. LNCS, vol. 8533, pp. 195–206. Springer, Heidelberg (2014)
3. Cheswick, W.: Rethinking passwords. Commun. ACM **56**, 40–44 (2013)
4. Florêncio, D., Herley, C., Van Oorschot, P.C.: Password portfolios and the finite-effort user: sustainably managing large numbers of accounts. In: Proceedings of the USENIX Security (2014)
5. Adams, A., Sasse, M.A., Lunt, P.: Making passwords secure and usable. In: Thimbleby, H., O'Conaill, B., Thomas, P.J. (eds.) People and Computers XII, pp. 1–19. Springer, London (1997)
6. Grawemeyer, B., Johnson, H.: Using and managing multiple passwords: a week to a view. Interact. Comput. **23**, 256–267 (2011)
7. Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: user attitudes and behaviors. In: Proceedings of the Sixth Symposium on Usable Privacy and Security, p. 2. ACM (2010)
8. Inglesant, P.G., Sasse, M.A.: The true cost of unusable password policies: password use in the wild. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 383–392. ACM (2010)
9. Boothroyd, V., Chiasson, S.: Writing down yourPassword: does it help? In: 2013 Eleventh Annual International Conference on Privacy, Security and Trust (PST), pp. 267–274. IEEE (2013)
10. Greene, K.K., Gallagher, M.A., Stanton, B.C., Lee, P.Y.: I can't type that! p@$$w0rd entry on mobile devices. In: Askoxylakis, I., Tryfonas, T. (eds.) HAS 2014. LNCS, vol. 8533, pp. 160–171. Springer, Heidelberg (2014)
11. Hayashi, E., Hong, J., Christin, N.: Security through a different kind of obscurity: evaluating distortion in graphical authentication schemes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2055–2064. ACM (2011)

12. Somayaji, A., Mould, D., Brown, C.: Towards narrative authentication: or, against boring authentication. In: Proceedings of the 2013 Workshop on New Security Paradigms Workshop, pp. 57–64. ACM (2013)
13. National Strategy for Trusted Identities in Cyberspace: Enhancing online choice, efficiency, security, and privacy (2011)
14. Marty, R.: Applied Security Visualization. Addison-Wesley, Upper Saddle River (2009)
15. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of the IEEE Symposium on Visual Languages, pp. 336–343. IEEE (1996)
16. Bergstrom, J.R., Frisch, S.A., Hawkins, D.C., Hackenbracht, J., Greene, K.K., Theofanos, M.F., Griepentrog, B.: Development of a scale to assess the linguistic and phonological difficulty of passwords. In: Rau, P.L.P. (ed.) CCD 2014. LNCS, vol. 8528, pp. 131–139. Springer, Heidelberg (2014)
17. von Zezschwitz, E., De Luca, A., Hussmann, H.: Honey, i shrunk the keys: influences of mobile devices on password composition and authentication performance. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, pp. 461–470. ACM (2014)
18. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**, 3–55 (2001)
19. Burr, W., Dodson, D., Perlner, R., Polk, W., Gupta, S., Nabbus, E.: Nist sp800-63-2-electronic authentication guideline. National Institute of Standards and Technology (2013)
20. Greene, K., Kelsey, J., Franklin, J.: Measuring the Usability and Security of Permuted Passwords on Mobile Platforms. National Institute of Standards and Technology Interagency Report (NISTIR) 8040 (2015)
21. Bostock, M., Ogievetsky, V., Heer, J.: D^3 data-driven documents. IEEE Trans. Vis. Comput. Graph. **17**, 2301–2309 (2011)
22. Tufte, E.R., Graves-Morris, P.: The Visual Display of Quantitative Information, vol. 2. Graphics Press, Cheshire (1983)
23. Tyler, C.W.: Human Symmetry Perception and its Computational Analysis. Psychology Press, Hove (2003)
24. Florêncio, D., Herley, C., Van Oorschot, P.C.: An administrators guide to internet password research. In: Proceedings of the USENIX LISA (2014)
25. Kelley, P.G., Komanduri, S., Mazurek, M.L., Shay, R., Vidas, T., Bauer, L., Christin, N., Cranor, L.F., Lopez, J.: Guess again (and again and again): measuring password strength by simulating password-cracking algorithms. In: 2012 IEEE Symposium on Security and Privacy (SP), pp. 523–537. IEEE (2012)
26. Weir, M., Aggarwal, S., Collins, M., Stern, H.: Testing metrics for password creation policies by attacking large sets of revealed passwords. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, pp. 162–175. ACM (2010)
27. Galbally, J., Coisel, I., Sanchez, I.: A probabilistic framework for improved password strength metrics. In: 2014 International Carnahan Conference on Security Technology (ICCST), pp. 1–6. IEEE (2014)