Scantegrity Responds to Rice Study on Usability of the Scantegrity II Voting System

December 28, 2014

Neal McBurnett, Richard T. Carback, David Chaum, Jeremy Clark, John Conway, Aleksander Essex, Paul S. Herrnson, Travis Mayberry, Stefan Popoveniuc, Ronald L. Rivest, Emily Shen, Alan T. Sherman, and Poorvi L. Vora

This note is a response to, and critique of, recent work by Acemyan, Kortum, Bryne, and Wallach regarding the usability of end-to-end verifiable voting systems, and in particular, to their analysis of the usability of the Scantegrity II voting system. Their work is given in a JETS paper [Ace14] and was presented at EVT/WOTE 2014; it was also described in an associated press release [Rut14]. We find that their study lacked an appropriate control voting system with which to compare effectiveness, and thus their conclusions regarding Scantegrity II are unsupported by the evidence they present. Furthermore, their conclusions are contradicted by the successful deployment experiences of Scantegrity II at Takoma Park.

Although their study reviewed three end-to-end verifiable voting systems (Pret A Voter, Scantegrity II, and Helios), we comment here only on their review of Scantegrity II.

They infer that these systems are "failing," [Ace14, p. 39] based in large part on their observation that in their experiments "only 58 percent of votes cast on tamper-resistant systems counted," [Rut14] with a graph showing approximately 50% of the Scantegrity II votes being counted, in their lab study with 37 voters [Ace14, p. 34].

The system they tested is *not* Scantegrity II as fielded in the 2009 municipal election at Takoma Park, and their study does *not* point out and explain the differences between the system they tested and the system used in Takoma Park in 2009. The Rice study imposed an extra hurdle on each voter: to scan the ballot and then *separately* drop it into a ballot box. Therefore, results from the Rice study should not be interpreted as properties of Scantegrity II as used in Takoma Park.

In addition, in the Rice study, investigators gave subjects instructions that were mismatched to the system they were using. Specifically, these instructions were based on written instructions to voters in the Takoma Park election, even though the Takoma Park system had an integrated scanner/ballot box but the Rice system separated the scanner from the ballot box. These mismatched instructions likely confused subjects.

Their paper reveals several misconceptions held by the Rice authors about how Scantegrity II works. They fail to recognize the *layered* nature of Scantegrity II in the design of their usability-testing protocols. More specifically: Scantegrity II adds a verification layer on top of a traditional optical scan voting layer; each voter may participate in the verification process by performing a few additional steps above and beyond what a traditional opscan voter would do.

The high failure rate observed by Acemyan *et al.* reflects aspects of their experiments that say nothing about the usability of the additional optional verification steps available to voters in Scantegrity II. Indeed, the failures they observe are all attributable to their mismatched voter instructions and their implementation of the underlying opscan layer, which separated scanner from ballot box. A separate ``control system'' that implemented the opscan layer only would have isolated the effects of the Scantegrity II verification layer. Unfortunately, their study did not include such a control system.

Furthermore, their paper does not identify the failure modes, but simply faults the usability of the entire system. The Rice study offers no evidence that the observed problems were due to the verification layer; by contrast, it appears that all failures occurred at the opscan layer implementation.

Observations from the Rice lab study run in sharp contradiction to what happened in the 2009 binding municipal election at Takoma Park [Car10a], when 100% of the 1728 votes were counted. These differences in ballot-casting rates are likely the result of differences between the system they tested (and their testing environment) and the system used in Takoma Park in 2009. Pleased with the performance of Scantegrity II, Takoma Park used Scantegrity II again in 2011 with similar results.

The Rice Study

The purpose of the Rice lab study was to investigate the usability of three end-to-end (e2e) verifiable voting systems: Scantegrity II, Helios, and Pret-a-Voter (PaV). Thirty-seven voters cast ballots using these three systems, voting as specified by a given intent sheet. The study randomized the order of the systems on which each voter voted. For each system, the study computed measures of effectiveness (ability to cast votes as specified), efficiency (time to vote), and voter satisfaction.

This note discusses only the Scantegrity II system and not Helios or PaV. Our comments about the layered nature of Scantegrity II probably do not apply to Helios because it does not share Scantegrity II's layered design.

Scantegrity II

Scantegrity II is a layered, add-on verification mechanism to optical scan voting where voters may opt-in to the verification process; voters who opt-out experience a voting procedure essentially the same as that for traditional optical scan voting. In addition to a standard "opscan layer" that provides basic voting capability, Scantegrity II adds a "verification layer" that provides the option of end-to-end verifiability.

Because of the layered design, the vote-casting process is familiar and standard to voters who have previously voted on opscan systems: (1) Check-in and obtain a paper ballot. (2) Fill in the ovals for one's choices. (3) Review the ballot to check that it correctly captures your intent. (4) Deposit the ballot in the scanner/ballot-box.

In addition to the standard opscan process, Scantegrity II provides additional optional mechanisms that are part of the end-to-end verification layer: (1) Record the previously-invisible confirmation code that appears when one marks on oval using a special pen. (2) Verify that these codes are correctly presented on the election web site. (3) File a complaint if an error or omission on the web site is observed.

Scantegrity automatically includes a mode of operation in which it is just a standard opscan system where voters do not need to use special pens. The scanners anyway only need to detect whether ovals are dark (they do not need to do OCR on the confirmation codes revealed). The special pens are necessary only if the voter wishes to use the end-to-end verification features; the special pens make the confirmation codes appear inside of each marked oval.

In addition to the many published papers on Scantegrity, the video from www.scantegrity.org plainly shows how Scantegrity II works, including the integral scanner/ballot box, which guarantees that all scanned ballots are cast into the ballot box, as implemented at Takoma Park.

Modes of Failure in the Rice Study

The Rice study shows a very low effectiveness rate (failure to complete) and implies that this low rate is due to the end-to-end verification features of Scantegrity II. But the problem lies with the implementation of the opscan layer, and not with the verification layer that Scantegrity adds.

Their paper does not say what the major failure mode is for "failure to complete" for Scantegrity II. But as explained by Acemyan at EVT/WOTE 2014, all but one of the uncounted ballots were invalidated because the voter failed to perform *both* of the following steps: (1) scan ballot; and (2) drop ballot into the ballot box. In Scantegrity II as used at Takoma Park in 2009, doing so would be impossible: the scanner is integrally attached to the ballot box; the only way for the ballot to enter the ballot box is through the scanner (see Figure 1).[1]

These two tasks are tasks from the opscan layer; they do not involve any end-to-end verification features. If a voter fails at performing any of these two tasks, then the opscan layer implementation is poor. Such a failure does not imply anything about the impact of end-to-end verification features on effectiveness.

Furthermore, the failure by a voter to scan his or her ballot is a recoverable error, since the ballot could be scanned later.

Opscan implementations usually have the scanner and ballot box as one integral unit, so that the common error mode found in the Rice study is not even possible. This error mode is commonly recognized as a problem. The standard best-practice solution, used both industry-wide for opscan systems, and in Scantegrity, is to have an integral scanner/ballot-box. A ballot is scanned and placed into the ballot box in one integral step. By contrast, in the Rice study, the implementation of the Scantegrity-inspired system had a separate scanner and ballot-box, and the voter had to take two separate actions. If the Rice study had evaluated the usability of any of the conventional precinct-based optical scan systems on the market in the same way (by separating the scanner from the ballot box and not providing procedural help to voters), we would have expected to have seen similar vote-casting failure rates.

Thus, the failure-to-complete error rate observed in the Rice study should be attributed to the poor design choice made for the opscan layer, contrary to best practice used in industry and unlike that used in Scantegrity II as fielded at Takoma Park.

Confusing mismatched instructions given to voters likely contributed to the high failure to complete ballot casting. In the Rice study, subjects were instructed, "To cast your vote, take your ballot to the scanner. Keep the card to verify your vote online after the polls close." [Ace14, p. 53]. Apparently, subjects were not instructed to deposit their ballots into the ballot box. By contrast, in the Takoma Park election, there was no need to instruct voters to deposit their ballots into the ballot box because the scanner dropped each ballot into the ballot box.

---

[1] This information also appears as fourth slide after the title from Sherman's 2011 presentation [She11] on Scantegrity III, and Slides 8 and 13 from Carback's 2010 presentation [Car10b] on the Takoma Park election.

In our 2009 mock election [She10] at Takoma Park preceding the municipal election, we implemented an early version of Scantegrity II in which the scanner was not attached to the ballot box. In that mock election, we avoided the type of confusion observed in the Rice study through election procedures (e.g., poll workers helped direct voters to deposit their ballots into the ballot box).

Also as explained by Acemyan at EVT/WOTE 2014, in the Rice study, one ballot was uncounted because the voter did not use the special pen with decoder ink. But in Scantegrity II, failing to use the decoder pen does not invalidate the ballot or prevent it from being properly counted; failing to use the special pen would only prevent the voter from verifying her ballot. The report is incorrect when it says that Scantegrity II requires "users to be an active part of the security process," or that it requires voters to use decoder pens or record the revealed invisible ink codes; the verification steps added by Scantegrity II are *optional*.

It is possible that the end-to-end features *could* have impacted the completion rates. For example, voters might have mistakenly cast their receipts rather than the ballots. Since receipts are present only in end-to-end verifiable voting systems, this error mode can be attributable to the end-to-end verification features. In Takoma Park, we did worry about this possibility, and made the verification cards a noticeably different size than the ballots so this would not happen. Or, voters might have been confused or distracted by the numbers appearing in the marked ovals, causing them to flee the poll site without voting. We have no evidence of this happening at Takoma Park, and the Rice study presents no evidence that this (or the other error mode just mentioned) was a problem for their voters.

Suggestions for Future Usability Studies

A well-designed usability study for Scantegrity II's end-to-end verification would have a "control system" that is *just* an opscan system implemented with the same design choices used by the implementation of Scantegrity II. We speculate that an opscan system implemented with a separate scanner and ballot boxs would not perform any better than did the Scantegrity-inspired system used in the Rice study, especially when voters are not given instructions to deposit the ballot into the ballot box. With such a control system, it would be possible to measure the impact on usability of the additional optional verification steps.

We appreciate the initial steps made in the Rice study regarding time-to-vote and on the usability of the online verification steps; these are aspects that are indeed affected by the end-to-end features of Scantegrity II. In particular, the online verification steps will be new to most voters, and the current implementations of online verification could almost certainly use further study and improvement.

Additional Issues with the Rice Study

We point out four other issues with the Rice study.

(1) In the Rice study, the investigators refused to answer any voter questions about the voting process. By contrast, in real elections, poll workers are willing to answer certain voter questions (*e.g.,* what to do next, or where to go next).

(2) In the Rice study, voters did not receive the same instructions that were provided to Takoma Park voters. In the Rice study, voters received only the instructions printed on the ballot. At Takoma Park, voters received a variety of instructions including written and video instructions while waiting in line, written instructions posted in the ballot marking stations, and written and video instruction provided

through television and newspapers prior to the election. And where part of the instructions were the same, they were inappropriate for the Rice system. When evaluating the usability of a system, one must consider both the system and its instructions.

(3) The report is also wrong to state that Scantegrity II "require[s] candidate order be randomized."

(4) Although the following point is relatively minor and unlikely to have affected ballot-casting rates, the Rice group incorrectly states that Scantegrity II receipts are stamped by an election official. In fact, Scantegrity II receipts were not stamped at Takoma Park in 2009. That Scantegrity II does not stamp receipts is a security advantage of Scantegrity II over some other systems. It is important that voters be able to make valid receipts without informing the election authority that they are in any different position from other voters concerning their ability to verify.

Conclusion

The Rice usability study shows a poor failure to complete (poor effectiveness) for their Scantegrity-inspired system, *but that poor effectiveness is due to poor choices made in the implementation at the opscan layer, confounded by confusing instructions to voters, and does not have anything to do with the end-to-end verification features that are the declared focus of their study*.

We predict that a pure opscan system implemented the same way as the opscan layer in their Scantegrity-inspired system, without appropriate instructions to voters, would show a nearly identical failure-to-cast rate. Conversely, we expect that a Scantegrity-like system implemented on top of a well-done opscan layer would show very high completion rates, as opscan systems typically do, and as Scantegrity did when fielded at Takoma Park. Achieving good usability of voting systems is crucial, and we applaud the Rice group for initiating the study of the usability of Scantegrity II and other end-to-end verifiable voting systems. However, readers of the Rice study should understand that the system they tested is not Scantegrity II as used at Takoma Park in 2009, and that the problems they observed appear *not* to be caused by the end-to-end verification features of Scantegrity II.

References

[Ace14] Claudia Z. Acemyan, Philip Kortum, Michael D. Byrne, Dan S. Wallach, "Usability of Voter Verifiable, End-to-end Voting Systems: Baseline Data for Helios, Prêt à Voter, and Scantegrity II," *USENIX Journal of Election Technology and Systems (JETS),* Volume 2, Number 3 (July 2014). www.usenix.org/jets/issues/0203

[Car10a] Richard T. Carback, Richard T., David Chaum, Jeremy Clark, John Conway, Aleksander Essex, Paul S. Herrnson, Travis Mayberry, Stefan Popoveniuc, Ronald L. Rivest, Emily Shen, Alan T. Sherman, and Poorvi L. Vora, ``Scantegrity II municipal election at Takoma Park: The first E2E binding governmental election with ballot privacy" in *Proceedings of USENIX Security 2010,* (Washington, DC, August 2010). http://www.usenix.org/events/sec10/tech/full_papers/Carback.pdf

[Car10b] Richard T. Carback, "Scantegrity II municipal election at Takoma Park: The first E2E binding governmental election with ballot privacy," slides from talk presented on August 12, 2010, at USENIX Security 2010 (Washington, DC). https://www.usenix.org/legacy/events/sec10/stream/carback/index.html

[Rut14] David Ruth, "Only 58 percent of votes cast on tamper-resistant systems counted," news release, Rice University (October 13, 2014). http://news.rice.edu/2014/10/13/only-58-percent-of-votes-cast-on-tamper-resistant-systems-counted/

[She10] Sherman, Alan T., Richard Carback, David Chaum, Jeremy Clark, John Conway, Aleksander Essex, Paul S. Herrnson, Travis Mayberry, Stefan Popoveniuc, Ronald L. Rivest, Anne Sergeant, Emily Shen, Bimal Sinha, and Poorvi Vora, "Scantegrity mock election at Takoma Park" in *Proceedings of the 4th International Conference on Electronic Voting 2010 (e-vote.cc 2010)*, (Lochau, Austria, July 2010).

[She11] Alan T. Sherman, EVT 2011. Alan T. Sherman, Russell A. Fink, Richard Carback, and David Chaum, "Scantegrity III: Automatic Trustworthy Receipts, Highlighting Over/Under Votes, and Full Voter Verifiability," slides from talk presented on August 8, 2011, at EVT/WOTE 2011 (San Francisco, CA). https://www.usenix.org/legacy/event/evtwote11/tech/slides/sherman.pdf

Figure 1. Scantegrity II ballot box / scanner. David Chaum stands beside a Scantegrity II ballot box with attached scanner, as used in the 2009 municipal election at Takoma Park, Maryland. It is impossible for a voter to drop a ballot into the ballot box without scanning it.