NYU TANDON SCHOOL OF ENGINEERING

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Visualizing Conversation: Development of Analytical Tools to Mitigate Cyber Fraud

Amanda L. Chiu, Jay Koven, Enrico Bertini

INTRODUCTION

Email-based scamming has been around for years and continues to succeed against unsuspecting targets. The ease, speed, and relative anonymity of emails makes it a powerful online tool for scammers. Fortunately, for forensic investigators, these web-based conversations leave a digital trail.

Understanding how scammers operate can provide insight on how to build better prevention methods and countermeasures against their attacks.

Extension of Beagle

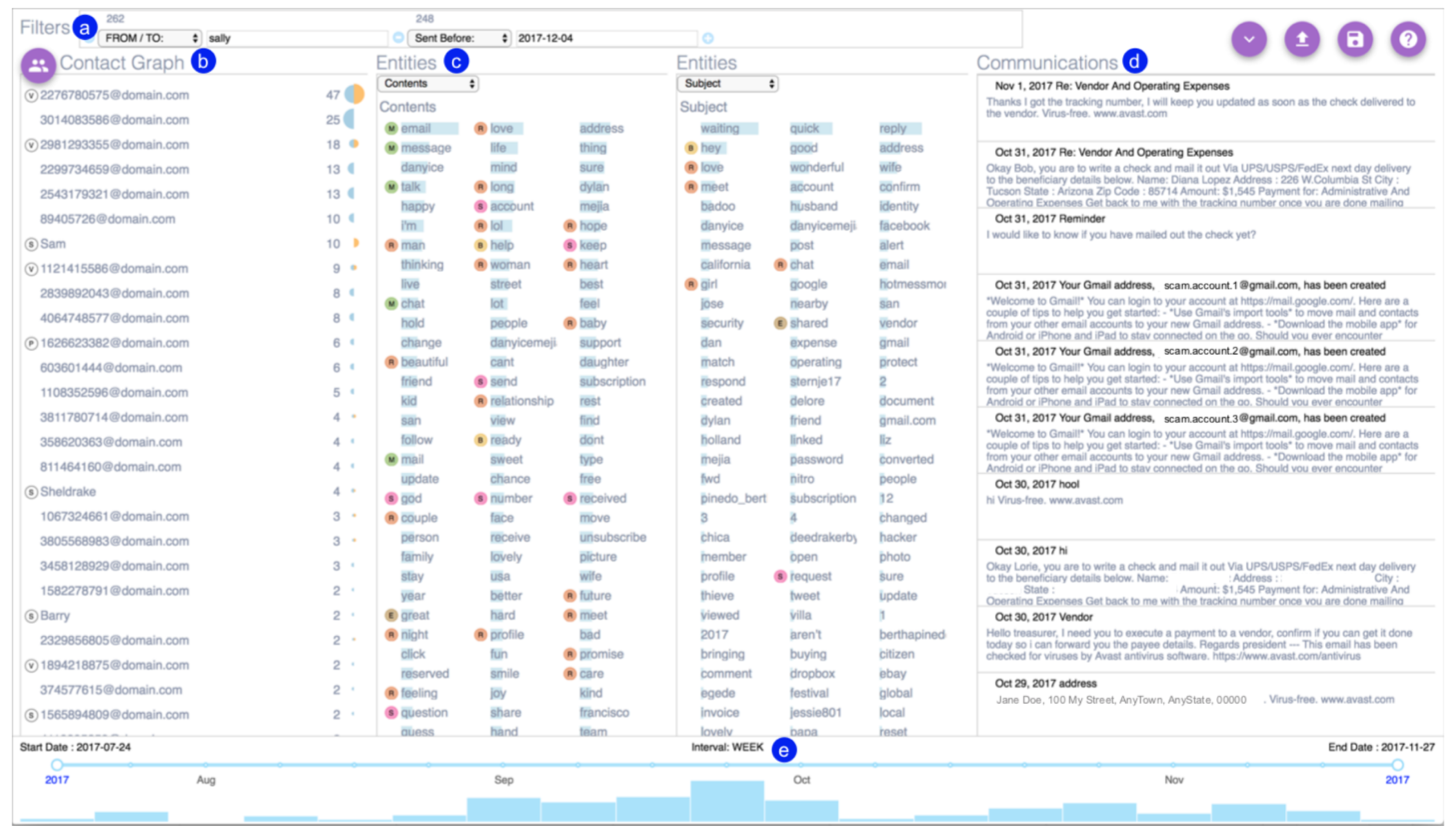

Figure 1. Forensic tool currently used by investigative teams to catch email fraud criminals [1]

Investigators gain insights from Beagle to identify new categories of email scam types, deconstruct large criminal networks, and stop ongoing fraud attempts. [1]

RESEARCH

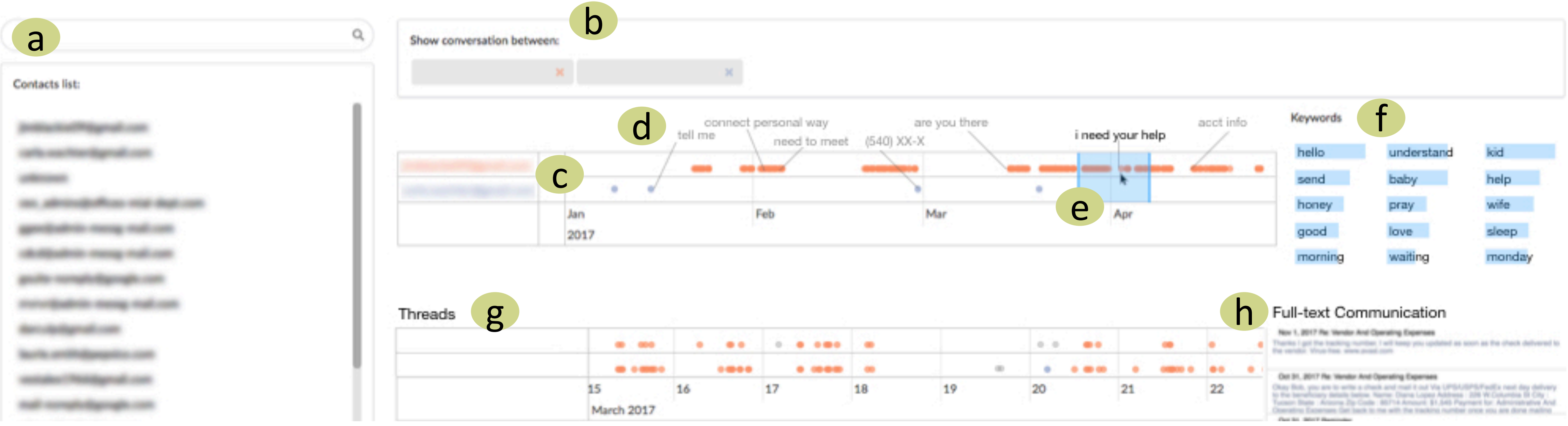

Figure 2. Visual analytics tool to investigate textual content and temporal patterns in email conversations

QUESTIONS TO ANSWER

- Can we spot a scammer based on their communication patterns? Can we distinguish different scammers?

VISUALIZATION AS A SOLUTION

(1) Current exploration of text-based data forces analysts to read and evaluate volumes of documents at the lowest level of abstraction – in this case, single email messages [1].

(2) Primary goal for investigators is to *understand* how scammers operate. Developing an interactive visual analytics tool allows humans to interact with the data and make intuitive decisions about what is significant/interesting.

EMAIL DATASET

In collaboration with a federal agency and forensic investigation team, we obtained a set of diverse conversations between scammers and their victims. The dataset contained 59,652 unique emails from 78 separate email accounts [1]

Target features of visualizing email:

(1) **WHO** emails addresses involved

(2) **WHAT** the content

(3) **WHEN** date & time sent

IMPLEMENTATION

Front-end: React-Redux application

Back-end: Node.js, GraphQL, Elasticsearch

LIST OF FEATURES **(a)** search or select email addresses of interest; **(b)** view query of email address(es) (correspondents); **(c)** scroll to zoom in/out temporal sequence of emails between correspondents; **(d)** text summarization of clusters of email; **(e)** brush over section of emails to see most relevant keywords within scope; **(f)** keywords extracted from emails; **(g)** temporal sequence of emails divided by thread; **(h)** full text of email messages

CONCLUSION

With the ability to study an overview of a text-based conversation, analysts can identify both general patterns and specific areas of interest for further investigation.

Applying visual analytics to explore this unique dataset has real-world impact. Advancing knowledge of how scammers operate allows investigators to identify and intercept scams faster, software services to integrate more proactive countermeasures, and the general public to be more aware of suspicious content.

FUTURE WORK

With classified patterns of scam activity, the detection of scams can be automated and flagged to bring to investigators' attention. Since email-based scamming involves social hacking to some degree, a visual representation of how sentiment (and how it evolves in longer conversations) can be another dimension to add.

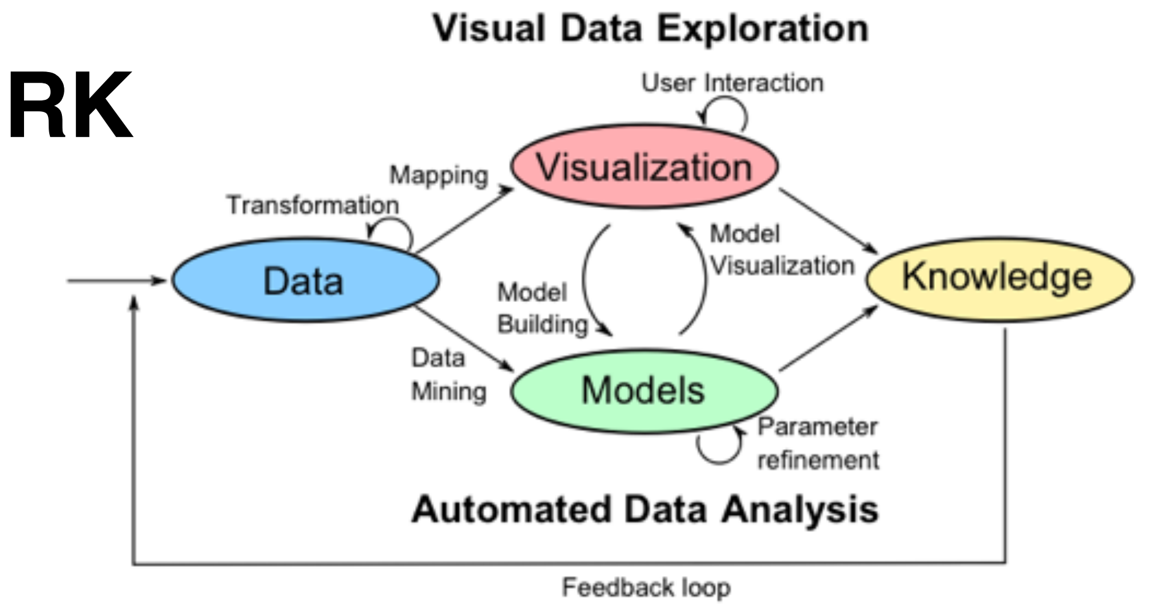

Figure 3. Visual Analytics Loop [2]

WORKS CITED

[1] J. Koven, C. Felix, H. Siadati, M. Jakobsson and E. Bertini. Lessons Learned Developing a Visual Analytics Solution for Investigative Analysis of Scamming Activities

[2]. Ed. Keim et al., "Mastering the Information Age, Solving Problems with Visual Analytics", 2010.