Using Web-Scale Graph Analytics to Counter Technical Support Scams

Jonathan Larson	Bryan Tower	Duane Hadfield	Darren Edge	Christopher White
Microsoft AI & Research	Microsoft AI & Research	Digital Crimes Unit	Microsoft AI & Research	Microsoft AI & Research
Microsoft	Microsoft	Microsoft	Microsoft	Microsoft
Silverdale, WA, USA	Silverdale, WA, USA	Redmond, WA, USA	Cambridge, UK	Redmond, WA, USA
jolarso@microsoft.com	brtower@microsoft.com	duhadf@microsoft.com	daedge@microsoft.com	chwh@microsoft.com

***Abstract*—Technical Support Scams delivered through malicious webpages and linked advertisements are an endemic problem on the web. To avoid detection, scammers systematically vary the URLs, web page designs, and phone numbers of scam delivery. We describe a web-scale pipeline powered by Cloud AI services that continuously detects and links evidence of such scams. By integrating this evidence in a graph structure and exposing it for forensic analysis in a user interface, we enable investigators and law enforcement partners to track the evolving scope and sophistication of scam operations. This approach automates and scales investigative tradecraft that contributed to major enforcement actions by the FTC in 2016. From 2016-2018, it helped reduce consumer exposure to such scams by 5 percent.**

Keywords—technical support scam, fraud, cybercrime, web-scale, graph analytics, visual analytics, forensic investigation

I. Introduction

This paper describes a collaboration between Microsoft Research and the Microsoft Digital Crimes Unit (DCU) to detect Technical Support Scams and disrupt the deceptive operations behind them [13]. Each month, Microsoft receives over 11,000 customer complaints about tech support scams [2], many of which are perpetrated by scammers claiming to be Microsoft or acting on Microsoft's behalf. As a threat to users of the web, the Technical Support Scam is notable for both its pervasiveness and persuasiveness: a 2016 global survey revealed that two thirds of internet users had encountered such a scam in the prior twelve months, one fifth had continued with a fraudulent transaction, and one tenth had lost money as a result [9].

Our approach to countering Technical Support Scams is threefold. First, we engineered a data pipeline for capturing potential scams in near real-time using distributed computation and machine learning (i.e., Cloud AI). This pipeline identifies 500-600 new scam webpages per day. Second, we use a network/graph formalism to model the pairwise relationships between the key elements of the scam – the URLs and domains hosting scam webpages (Fig. 1), the visual signatures of those webpages, and the obfuscated toll-free phone numbers embedded within them. Third, we developed forensic interfaces for interactive analysis of the resulting evidence networks, allowing users to view images of the captured scam webpages, to view patterns of behavior over time, and to search and filter based on associated metadata. Use of this tool by DCU analysts supports Microsoft's collaboration with international law enforcement agencies to counter the global threat of Technical Support Scams. While the design of the tool was based on the investigative tradecraft that contributed to 16 major enforcement actions taken by the FTC against scam organizations and their owners in 2016 [7], its subsequent deployment from 2016-2018 has led to further actions by Indian law enforcement on Delhi-based call centers and helped to drive a global 5-percent decline in consumer exposure to ad-based Technical Support Scams [2].

Fig. 1. Technical Support Scam with text obfuscated by image embedding.

II. Adversarial Evolution of Technical Support Scams

Technical Support Scams began to appear in 2008 as a low-tech evolution of fake security product tele-sales [11]. The prospective victim is cold-called, told there is a serious problem with their computer (e.g., a virus, trojan, or other form of malware), and that the technician needs remote access to the computer to diagnose and repair the underlying issue. The scammers use social engineering techniques both before and during the remote access session to convince the target that their computer has been compromised, often by misrepresenting the routine output of system tools as evidence of non-existent problems (e.g., revealing "critical" errors in `eventvwr`, "rogue" connections in `netstat`, and "malicious" files listed by utilities including `prefetch`, `inf`, and `dir` [11][15][16]). Once the target appears convinced, high-pressure sales tactics are used to persuade them to pay for unnecessary repair services, subscription-based service plans, and anti-virus software and

other products or services [7]. This social engineering remains at the core of the scam, with typical payment requests of several hundred US dollars. Even if the user complies, they remain vulnerable to overcharging, covert installation of malware, and exfiltration of financial or other sensitive information [15]. On the other hand, failure to pay can lead to scammers remotely setting passwords that lock users out of their machines [16], either as punishment or for escalation to a ransom demand.

A 2014 FBI Public Service Announcement described an emerging form of Tech Support Scam delivery based on a combination of "scareware" and "malvertising" [12]. Scammers would first purchase domains imitating familiar security brands, or containing scare terms related to malware, hacking, and infection. Webpages hosted on these domains would inform the user of a problem with their computer and provide a toll-free phone number on which the user could receive immediate assistance. The scammers would then purchase ads linking to these webpages and distribute them across the web via ad platforms. In a further evolution, these ads began to use intrusive JavaScript techniques to navigate to scam pages automatically on loading, before generating pop-up notifications that the user could not easily dismiss [15]. As the styling of these webpages often resembled system error messages, the deception could be quite convincing. Some even including count-down clocks or scary audio messages to add a sense of urgency [7]. Compared to cold-calling, users dialing the toll-free number had self-selected and were therefore more likely to be susceptible to victimization. Moreover, such ad-based prospecting could be deployed at a significantly larger scale than cold-calling.

As Technical Support Scams have grown in public awareness and research interest, scammers have learned to demand up-front payment rather than risk "scambaiting" [1] or systematic study of scam techniques [15][16]. Growing awareness by scammers about the risks of including phone numbers in plain text (e.g., [15]) has also led to range of countermeasures, including use of live-chat [16], SEO for listing fake technical support websites in organic search results [17], on-the-fly delivery of time-limited and context-dependent phone numbers [15], and the obfuscation of phone numbers and scam text within images on the hosting webpage [10]. It is this latter phenomenon that we address in this paper.

III. Related work in cybercrime investigation

Phone numbers often play a key role in enabling communication between cybercriminal and victim. They also provide a proxy representation of the criminal when attempting to detect and connect evidence from disparate sources. For example, the infamous Koobface malware gang were exposed because of insecure source code backups that contained gang aliases and phone numbers for reporting daily revenue statistics [5]. Cross-referencing these against online forums and social media helped reveal the true identities of the gang members.

In the investigation of Nigerian/419 Email Scams, bipartite graphs of phone numbers and the email addresses containing them have provided a valuable abstraction with which to understand the structure, size, and behavior of scammer communities [4]. In these scams, phone numbers tend to be a more stable representation of a scammer than disposable (and spam-filterable) email addresses, but both phone numbers and email addresses showed significant reuse. A similar pattern of reuse has also been observed in Japanese One Click Fraud across phone numbers, domain names, and bank account numbers [3].

The most comprehensive study of Technical Support Scams to date [15] is also based on analysis of the bipartite graph formed by the toll-free phone numbers contained in the text of fraudulent webpages and the domains hosting them. The same paper describes the ROBOVIC crawler for automatically discovering and linking instances of Tech Support malvertising and the results obtained over a 250-day collection period. From 5 million scanned domains, it recorded 22k scam-related URLs across 9k unique domains, referencing 1.5k phone numbers. Analysis revealed that although the average lifetime of a scam URL was 11 days, 43% of URLs were only valid for less than three days. The paper also analyzes the scam ecosystem and its economics, as well as the results from an observational study interacting with 60 scammers. The study yielded an average service price of $291 USD, call duration of 17 minutes, and estimated call-center size of 11 operators.

IV. Countering technical support scams with graphs

A. AI pipeline for detecting and linking scams at web-scale

Microsoft encourages consumers to report any encounters with Technical Support Scams using the online form at `www.microsoft.com/reportascam` [10], which receives over 11k complaints per month. In 2017, complaints totaled 153k from 183 countries, up 24% on the previous year. Details reported in these complaints are all useful seeds for investigation, but pursuing each case individually is beyond the scope of human investigation. Shifting from a reactive, complaint-driven approach to a proactive, data-driven approach was crucial for tackling scams as they evolve on a global basis.

The heart of our detection pipeline is a reactive web scraper that captures images of scam webpages as they are detected from a variety of near real-time signals. Speed of evidence capture is critical as many scam URLs are live for just a few hours. This pipeline identifies around 150k suspicious URLs per day, together hosting around 100k new images potentially containing text relating to a Technical Support Scam. A range of Microsoft Cloud AI services are then used to confirm the likely presence of a scam, including: (a) Microsoft CNTK image classifiers trained to detect images of scam webpages; (b) Microsoft Cognitive Services Optical Character Recognition (OCR) to extract image text for further analysis; and (c) text classifiers trained to detect scam keywords and phrases. Custom phone number extraction and the generation of visual signatures from the images of captured websites complete the pipeline, allowing linking of scam webpages by shared design and/or phone number. This process yields around 500-600 new instances of image-obfuscated Technical Support Scams per day.

As in the related work, we use a graph formalism to model the connections between the URLs, domains, and visual signatures of scam webpages, and the embedded phone numbers extracted from them. Our work is differentiated from prior work in terms of both the scale of scam detection and the interfaces we provide for interactive exploration and forensic analysis.

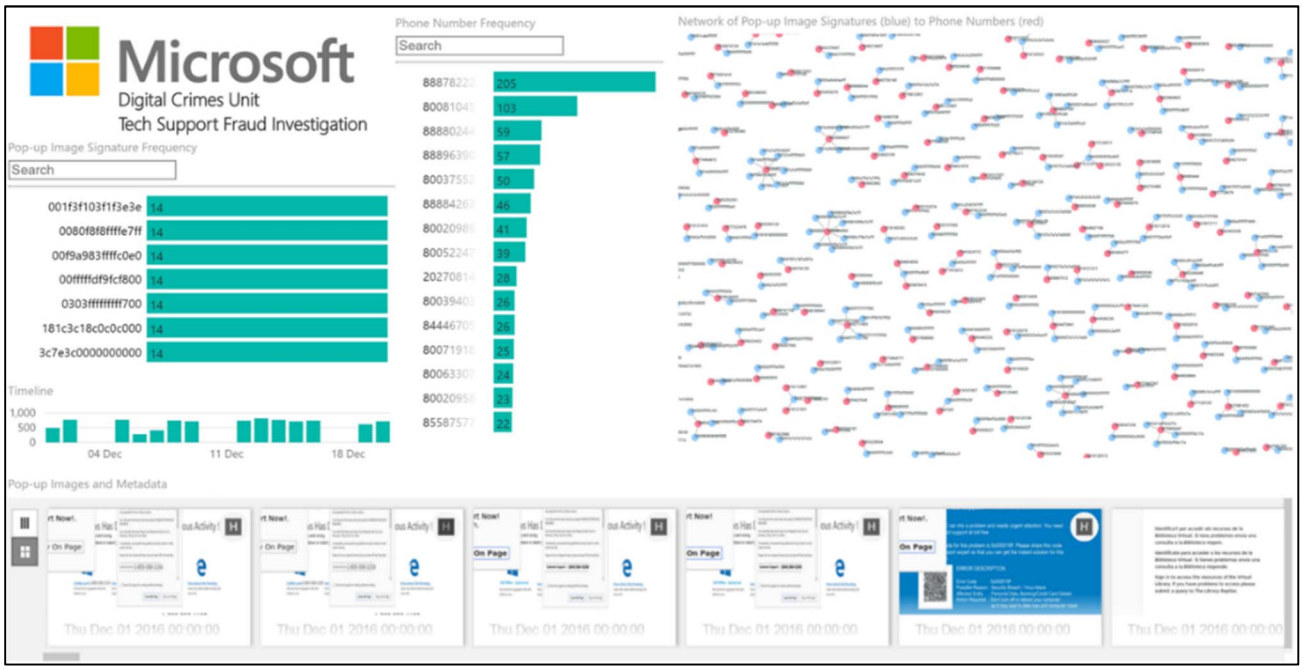

Fig. 2. Interface for exploring images of scam webpages, linked and extracted metadata, and image-to-phone-number networks in Microsoft Power BI.

B. Visual analytics interface for exploring scam networks

Our AI pipeline mines evidence of potential scams with high confidence, but it remains important for analysts to:

1. get an overview of activity over various timescales;
2. prioritize scam operations for forensic investigation;
3. filter based on metadata of interest (e.g., domain);
4. verify the metadata extracted from the raw images;
5. search for evidence associated with an external lead;
6. share findings with colleagues and law enforcement.

To help analysts achieve these goals, we used Microsoft Power BI to develop a collection of user interfaces to the data generated by the Technical Support Scam pipeline. Power BI enables data scientists to build data models and interactive dashboards over linked data sources, and share these with end-users for interactive summarization, exploration, and querying of the resulting data streams. An overview of how we have extended Power BI with representations that enable analysis of unstructured and graph-structured data can be found in [6].

Each of our interfaces addressed the above questions by focusing on a different element of the data model, including the raw images and their extracted meta-data; the frequencies of extracted image signatures, phone numbers, and URLs; the geographic distribution of IP addresses; and the characteristics of the resulting networks. Fig. 2 shows the overview interface listing visual signatures and phone numbers by frequency, the frequency of new scam detection over time, the raw pop-up images, and the bipartite network of visual signatures (blue nodes) linked to phone numbers (red nodes). Selecting any element of any visual representation (e.g., a phone number from the frequency-ranked list) filters the other representations to only show data linked to that selection (e.g., the images, signatures, timeline, and network of the selected phone number). Selecting the thumbnail of a webpage also expands it to show all linked and extracted meta-data.

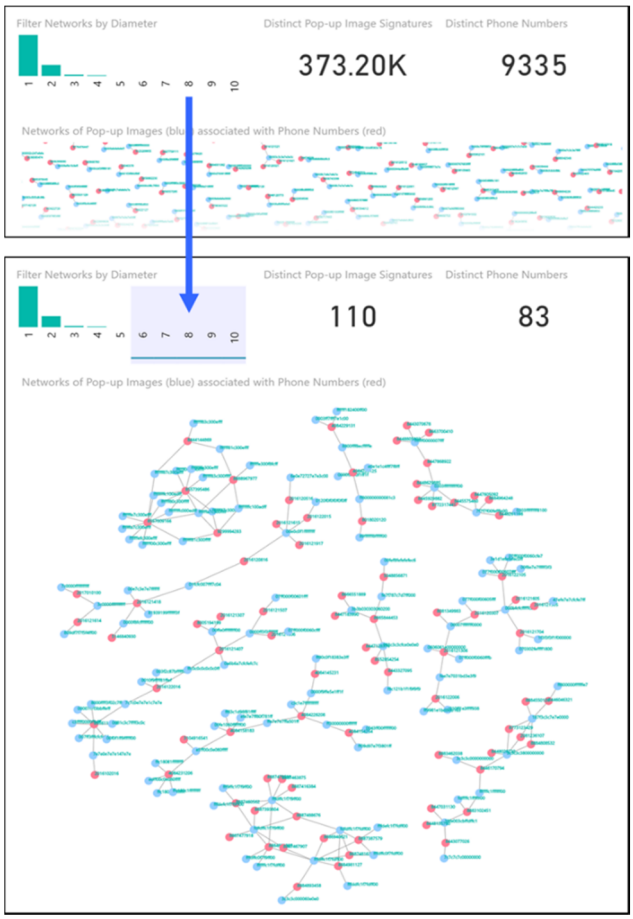

Fig. 3. Using network diameter to prioritize investigation of the most sophisticated scam operations. Selecting networks of diameter 6-10 filters the data to reveal patterns of deception arising from reuse of webpage designs and phone numbers across scams, while dramatically reducing the number of leads to investigate (373,200→110 signatures, 9335→83 phone numbers).

A graph metric that has proven particularly useful in prioritizing investigative efforts has been the diameter of the webpage signature to phone number subgraphs (Fig. 3), since most connected components in the network are very small. Each link in one of these subgraphs represents an attempt to evade detection by systematically varying the phone number or webpage design. Larger subgraphs represent more sophisticated operations and hence more organized instances of cybercrime. Analysts can select each node of a subgraph in turn to examine the webpage images and meta-data behind the connections, capturing visual evidence for sharing with law enforcement.

V. Discussion

The Tech Support Scam detection pipeline and forensic analysis interface described in the previous section are used on an ongoing basis by the Microsoft Digital Crimes Unit. It supports industry-wide technology collaboration with companies including Apple, Dell, Yahoo, and HP, as well as civil and criminal case referrals to law enforcement agencies worldwide. The impact of these technologies has been to automate the investigative tradecraft that has already contributed to major law enforcement actions by the US Federal Trade Commission [7]. Under "Operation Tech Trap", the FTC and its federal, state, and international partners (including Microsoft) have helped to build cases against some of the worst perpetrators of Technical Support fraud [7]. This round of 16 actions included complaints, settlements, indictments and guilty pleas for organizations and individuals engaged in tech support fraud globally. To give a sense of the harm caused by a single organization named in these actions, "Client Care Experts" defrauded more than 40,000 victims in excess of $25 million USD over a three-year period [10]. Together, the automation of scam discovery and the visual representation and analysis of linked scam activity time allow for a significant increase in the volume of addressable cases. The impact of this transformation, supplemented by consumer education, has been a 5-percent decline in consumer exposure to Technical Support Scams worldwide (from 68% to 63%), coupled with fewer financial losses (from 9% to 6%). It has also led to recent raids on ten call-centers in Delhi, with 24 owners and team-leaders arrested [2].

We are also actively involved in enhancing engineering solutions that aim to prevent scam delivery through ad platforms and the Edge web browser. In 2017, Microsoft's Bing ads organization rejected 650 million bad ads – a fivefold increase over 2016 [8]. 25 million of these ads related to Technical Support Scams, resulting in 3,500 scam organizations and 240k bad advertisements being permanently banned by the Bing ads platform. By integrating data from our own scam detection pipeline, we hope to increase this number for 2018 and beyond. Bad ads cannot be delivered if the associated URL is blocked by the web browser or operating system, which is why we are also working with Windows Defender SmartScreen [14] to increase their URL blocking ability for scam-hosting URLs.

VI. Conclusion

Technical Support Scams have come a long way since cold-calls unamenable to an "anti-scammer" solution [11]. Scam delivery has evolved into image-based webpages that are linked to by malicious ads and which use intrusive JavaScript to lock the web browser. By combining AI pipelines, graph analytics, and visual interfaces to linked evidence, we have taken a web-scale approach to detecting and protecting against these scams.

VII. Acknowledgments

We thank Donal Keating, Michael McDonald, Zoe Krumm, and Courtney Gregoire of the Microsoft Digital Crimes Unit.

References

[1] J. Brodkin, "'Can you fix my Windows 95 computer?': How to troll a tech support scammer", 10 Oct. 2012. Online, accessed 5 Nov. 2018: https://arstechnica.com/information-technology/2012/10/can-you-fix-my-windows-95-computer-how-to-troll-a-tech-support-scammer/

[2] A. Chansanchai, "Online scammers cost time and money. Here's how to fight back", 15 Oct. 2018. Online, accessed 5 Nov. 2018: https://news.microsoft.com/on-the-issues/2018/10/15/online-scammers-cost-time-and-money-heres-how-to-fight-back/

[3] N. Christin, S.S. Yanagihara, and K. Kamataki, "Dissecting one click frauds", 2010. ACM Conf. on Computer and Comm. Security, 15-26.

[4] A. Costin, J. Isacenkova, M. Balduzzi, A. Francillon, and D. Balzarotti., "The role of phone numbers in understanding cyber-crime schemes", 2013. IEEE Conf. on Privacy, Security and Trust, 213-220.

[5] J. Drömer and D. Kollberg, "The Koobface malware gang exposed", Jan. 2012. Online, accessed 5 Nov. 2018: https://www.sophos.com/en-us/medialibrary/pdfs/other/sophoskoobfacearticle_rev_na.pdf

[6] D. Edge, J. Larson, and C. White, "Bringing AI to BI: enabling visual analytics of unstructured data in a modern Business Intelligence platform.", 2018. ACM CHI 2018 Conference Extended Abstracts.

[7] Federal Trade Commission, "FTC and Federal, State and International Partners Announce Major Crackdown on Tech Support Scams", 12 May 2017. Online, accessed 5 Nov. 2018: https://www.ftc.gov/news-events/press-releases/2017/05/ftc-federal-state-international-partners-announce-major-crackdown

[8] N. Garg, "Bing ads Ad quality year in review 2017", 17 Apr. 2018. Online, accessed 5 Nov. 2018: https://advertise.bingads.microsoft.com/en-gb/blog/post/april-2018/ad-quality-year-in-review-2017

[9] C. Gregoire, "Tech support scams are a growing problem", 17 Oct. 2016. Online, accessed 5 Nov. 2018: https://blogs.microsoft.com/on-the-issues/2016/10/17/tech-support-scams-growing-problem

[10] C. Gregoire, "The fight against tech support scams", 18 May 2017. Online, accessed 5 Nov. 2018: https://blogs.microsoft.com/on-the-issues/2017/05/18/fight-tech-support-scams/

[11] D. Harley, M. Grooten, S. Burn, and C. Johnston, "My PC has 32,539 errors: how telephone support scams really work", 2012. Virus Bulletin.

[12] Internet Crime Complaint Center, "New Twist to the Telephone Tech Support Scam", 13 Nov. 2014. Online, accessed 5 Nov. 2018: https://www.ic3.gov/media/2014/141113.aspx

[13] A. Linn, "How Microsoft used AI to help crack down on tech support scams worldwide", 15 Jun. 2017. Online, accessed 5 Nov. 2018: https://blogs.microsoft.com/ai/microsoft-used-ai-help-crack-tech-support-scams-worldwide/

[14] Microsoft, "Windows Defender SmartScreen", 27 Jul. 2017. Online, accessed 5 Nov. 2018: https://docs.microsoft.com/en-us/windows/security/threat-protection/windows-defender-smartscreen/windows-defender-smartscreen-overview

[15] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial one for scam: analyzing and detecting technical support scams", 2017. NDSS 2017.

[16] R. Sampsa and V. Leppänen, "'You have a potential hacker's infection': a study on technical support scams", 2017. IEEE Computer and Information Technology (CIT), 197-203.

[17] B. Srinivasan, A. Kountouras, N. Miramirkhani, M. Alam, N. Nikiforakis, M. Antonakakis, and M. Ahamad, "Exposing search and advertisement abuse tactics and infrastructure of technical support scammers", 2018. WWW 2018, 319-328.