JDST., vol. 4, no. 3, p37-p52, 2021
https://doi.org/10.46713/jdst.004.03

www.jdst.eu, e-ISSN 2534-9813

Research Article

Interactive Stereoscopically Perceivable Multidimensional Data Visualizations for Cybersecurity

Kaur Kullman [a], Don Engel [a]

[a] *University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, US*
https://csst.umbc.edu

ABSTRACT

Interactive Data Visualizations (IDV) can be useful for cybersecurity subject matter experts (CSMEs) while they are exploring new data or investigating familiar datasets for anomalies, correlating events, etc. For an IDV to be useful to a CSME, interaction with that visualization should be simple and intuitive (free of additional mental tasks) and the visualization's layout must map to a CSME's understanding. While CSMEs may learn to interpret visualizations created by others, they should be encouraged to visualize their datasets in ways that best reflect their own ways of thinking. Developing their own visual schemes makes optimal use of both the data analysis tools and human visual cognition.

In this article, we focus on a currently available interactive stereoscopically perceivable multidimensional data visualization solution, as such tools could provide CSMEs with better perception of their data compared to interpreting IDV on flat media (whether visualized as 2D or 3D structures).

ARTICLE INFO

RECEIVED: 09 Oct 2021
REVISED: 10 Nov 2021
ACCEPTED: 30 Nov 2021
ONLINE: 12 DEC 2021

KEYWORDS

Stereoscopically Perceivable, Immersive, Data Visualization, Interactive Data Visualization, Cybersecurity, Virtual Reality, Augmented Reality, Mixed Reality, Interactive Stereoscopically Perceivable Multidimensional Data Visualization

✉ Corresponding Author: E-mail: jdst@coda.ee

I. Overview

As commercially available virtual [1], augmented [2] and mixed reality [3] (VR, AR, MR; collectively "xR") devices have become significantly more performant over the last decade, there has been a commensurate growth in interest in using these tools for three-dimensional data visualizations. Most of this interest has been in (geo)spatial data visualization [4] [5] [6], i.e., the visualization of imaginary, proposed, or real physical environments with overlayed textual information [6] [7]. Researchers and practitioners have focused less on how to represent non-(geo)spatial data using stereoscopically perceivable multidimensional data visualizations (SPMDV).

As with all other types of interactive data visualizations (IDV), interactive SPMDV (ISPMDV) should be created with or by subject matter experts (SMEs) in order to ensure that these creations will indeed serve their intended audience well [8]. To enable cybersecurity SMEs (CSMEs) to create useful stereoscopically perceivable IDVs (SPIDVs), these CSMEs need (at least):

1) An easy-to-follow method identifying what to visualize.
2) Easy-to-configure tools for creating the visualizations proposed in (1).
3) Tools which enable ingesting data from its source (e.g., SIEM, log correlation) into the visualization created in (2).

Although (2) and (3) may be combined into one tool, the objectives of (2) and (3) are distinct; (2) focuses on data visualization (in an xR headset), while (3) deals with "translating" ingested data from its source to a preferred format that would be suitable for (2).

In this paper we will give an overview of such a method and combined tool.

II. What to Visualize and How

ISPMDV can be considered an "add-on" to SPMDV, which in turn derives from multidimensional data visualizations (MDV). While MDV on flat screens is a well-researched topic [9] [10] [11] [12], SPMDV has received broader public attention only gradually during the past ten years [13] [14] [5] [15], with the emergence of VR and MR headsets that are good enough to have enabled researchers [16] [17] [18] and practitioners [19] [20] [21] to explore their capabilities for data visualization.

Being fundamentally spatial in nature, geospatial data visualization [4] and graphs [19] have relatively straightforward implementations in SPMDV. Given that cybersecurity data is not intrinsically spatial, then for which CSME tasks would SPMDV visualizations be useful? Or rather, what kind of SPMDVs and ISPMDVs would best suit the CSME tasks, and how might these be designed? CSMEs rely on large datasets, so it stands to reason that the full use of a third dimension afforded by xR will be useful in making more data visually discernable without relying on cluttering cues like shading, occlusion, and perspective (as is needed for MDV on flat screens). Model Mapping Method for Cybersecurity (M4C) [8] is one method

that can be used to design SPMDV while enabling the CSMEs to serve as both the designers and consumers of their own visualizations. In M4C, visualizations of networked entities (e.g., computers) are positioned according to their logical (but not necessarily their physical) topology, with the resulting 3D structure(s) matching a CSME's understanding of a network's expected topology. While other methods exist [22] [23], the essence of M4C is to extract the understanding of a dataset that CSMEs use for their tasks. That extracted understanding is then used in the process of creating such an SPMDV or ISPMDV for these CSMEs, that would enable them to further explore their dataset in ways that are aligned to their internalized understanding of that dataset.

It is important to note that these visualizations cannot be created independently of the CSMEs who would be using these visualizations for their tasks. An ISPMDV may look like a fancy scene pulled from science fiction to third parties, but it must be useful for its users; otherwise, these visualizations are not worth the cost of electricity that is needed to power the equipment that is running them. For verifying effectiveness, metrics established for the evaluation of ISPMDV should be used [24].

To explain the idea of creating data-specific 3D layouts further, let's take a simplistic computer network topology and lay it out onto a three-dimensional data-shape based on the IP addresses that are used by the entities in that network.

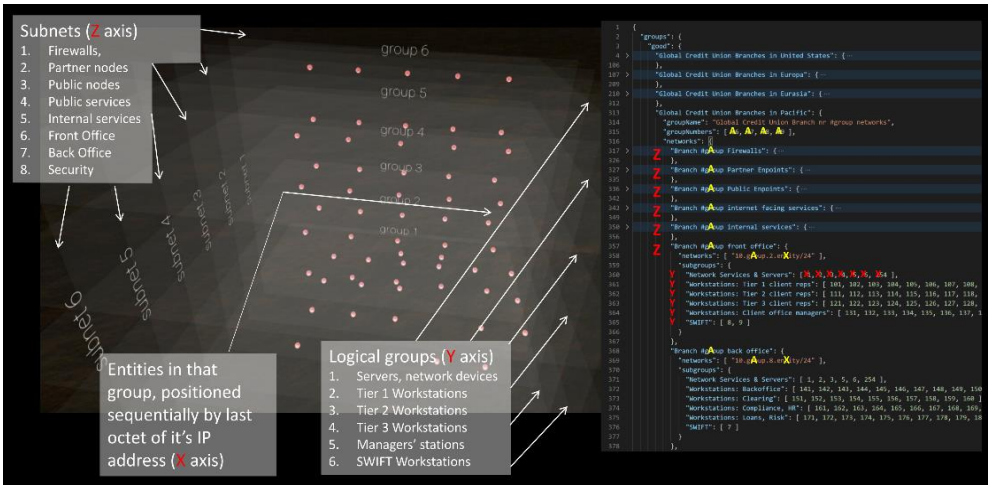

Figure 1. Networked entities arranged in a cube shape based on their functional topology; respective configuration of such data-shape shown on the right (from [25]).

A demo (mock-up) dataset consisting of an imaginary credit union's corporate and branch networks traffic (and its logical topology) is distributed with Virtual Data Explorer (VDE) [26]. We will discuss VDE in more detail in the next section, but for now, we focus on how a CSME would define a 3D layout of a dataset using VDE.

The left side of Figure 1 depicts a simplified version of the three-dimensional structure of that imaginary credit union's branch network (data-shape), with the

reddish spheres marking entities on the network. On the right side of Figure 1 is the snippet from the configuration file which VDE uses to map ingested data (in this case, observed and filtered network traffic) to spatial structures. From this vantage point, we can use the common XYZ axes to describe the Figure 1 data-shape, but it doesn't make much sense to stick with the XYZ thinking in complex constellations.

In the configuration example on the right of Figure 1:

1) the red letters Y and Z refer to the spatial positions of entities in that group in the data-shape, respectively, on the Y and Z axes,
2) the red X refers to the sequential position (on the X axis inside that group (Y)) of an entity with a matching IP address,
3) the yellow 'A's refer to group numbers (in this example, matching with the second octet of entity's IP address) and
4) yellow 'X's refer to the IP-address' last octets.

Note that the "networks" descriptor in the configuration (line 358), contains two variables – "group" and "entity". Subgroup members are mapped to the "entity" part of it (the last octet of the IP address in this case, yellow or red X), while the "group" number (line 315) refers to both the branch number (lines 317, 327, 336, etc.) and the second octet of an entity's IP address template (line 358).

Determining which subgroup contains a given entity, and determining which parameter(s) serve as the basis for this decision, is completely up to the CSME author of the visualization. This grouping is up to their perception, based on their understanding of the dataset. In this example case, these are the business functions of involved devices and their groups.

Such data-shapes representing groups of entities can then be positioned into constellations according to the CSME's understanding of the dataset that is being visualized. For example, the prerequisite knowledge needed from a CSME to create such a constellation containing a set of proposed data-shapes for depicting a computer networks functional topology would be to:

a) Understand the principles of how a computer network functions; specifically, how is such a network set up in the environment that the author of this visualization (the CSME) needs to understand.

b) Understanding the logical grouping of networked entities and their topology, but also networked entities and stakeholders' goals (e.g., corporate, employees, external {friendly, neutral, malicious} actors, etc.).

c) Understanding the expected behavior of the above actors and how it might be reflected in network data.

d) What indicators to look for, how to validate the findings, how to act with that combined knowledge.

Please refer to [8] for further information on how to design such data-shapes with CSMEs for CSMEs.

III. ISPMDV Examples

Although there are use cases for SPMDVs without user interaction, the implementation of even simple interactions can significantly accelerate a user's familiarization with visualized data. More importantly, the ability for the user to interact, select, and alter the selection of visualized (or augmented) data via queries greatly enhances the CSME's ability to learn to interpret the visualization.

It has been shown that first-time users tend to intuitively reach out to the data representations, as if to verify the existence of these artificial objects hanging in front of them [27]. Haptic feedback (using controllers), auditory cues, and realistic shaders further enhance users' immersion in an ISPMDV environment.

To create and customize ISPMDVs for CSMEs with CSMEs, Virtual Data Explorer (VDE) software was created with US Army Research Lab support [27]. VDE has three components:

1) A backend, which interprets the configurations of your network topology (json) and maps the ingested data according to that config into groups (of groups (of groups (of groups))) of entities.
2) A browser plugin, which helps in feeding the data from a Moloch, SIEM or custom log correlation tool as appropriate (say, after running a query), via a WebSocket to the VDE backend.
3) A headset (Magic Leap, Oculus, MS Mixed Reality, HTC Vive, HoloLens 2), which gets the set of groups from the backend and positions these for the viewer according to the selected layout configuration (json).

Unity 3D is used to create the software responsible for the ISPMDV in the headsets, C# is used for the backend, and JavaScript is used for the browser plugin.

Due to the medium on which you are reading this paper being flat (a screen or a physical piece of paper), it is impossible to convey here the "spatialness" of ISPMDV with figures. The next-best option to observing ISPMDV examples directly in xR are first-person videos of users interacting with an ISPMDV; for these, please visit coda.ee/JDST. Below are a few screenshots from video captures of VR and MR sessions, where the user either explores or interacts with an ISPMDV.

A. NATO CCDCOE CDX Locked Shields

To test the utility of ISPMDV when encoding non-spatial data, networked entities that were found to be present in NATO CCDCOE Locked Shields Cyber Defence eXercise (LS) [28] network traffic were spatially positioned as semi-transparent spheres, according to entities' positions in that (Blue Team's) network's functional topology, and, more importantly, entities' affiliation with logical groups present in LS networks. Logical groups could be distinguished by their members' functionality (e.g., SCADA components), purpose (e.g., DMZ servers), risk exposure, operating system, etc. (see Figure 2). This resulted in custom 3D data-shapes that were combined into a constellation (a VDE layout) representing a larger whole of the LS network(s). Constellations shown in Figures 3-5 depict LS network traffic with

VDE (in VR or MR), while Figures 6 and 7 use a slightly more primitive approach with OpenGraphiti (in VR).

LS traffic was chosen for testing purposes due to the complexity of its network and because the first author, having been involved with LS since 2010, has a deep understanding of that network's topology and its components' expected behavior.

Network topology (positions of groups of groups of groups of nodes), displayed as spheres and cubes, is overlayed with network traffic, visualized as lines (edges) rendered as connecting the nodes that were observed to have initiated or received network connections (sessions) during a time window. Which edges are visible, their opacity, and their coloring depend on the results of the underlying query that was executed once the user defined the time window and other query parameters (e.g. ASN, IP range, ports used, and amount of data transferred).

Figure 2: VR view of Locked Shields 18 Partner Run, focus on 11th Blue Team's networks (refer to config example, from line 313 onward, in Figure 1): subnets' labels are visible above the blades ("Blue Team 11 Firewalls," "Threod drone ground station …," etc.), that contain groups of entities ("OSX in INT network," "Services in INT network," etc.), that, in turn, contain spheres representing individual entities (IP address) of that group.

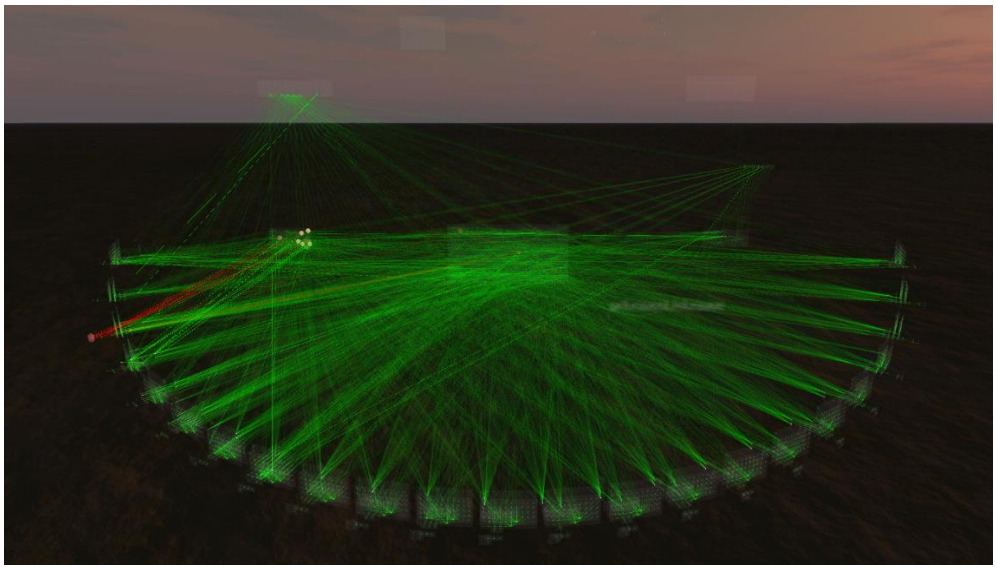

Figure 3: VR view of Locked Shields 18 Partner Run network topology and network traffic using VDE, displaying an overall view of the meta-shape: a data-shape consisting of multiple data-shapes. Red edges represent selected connections between Blue Team 3 device and Red Team nodes. A detailed description of this layout can be found in [21].

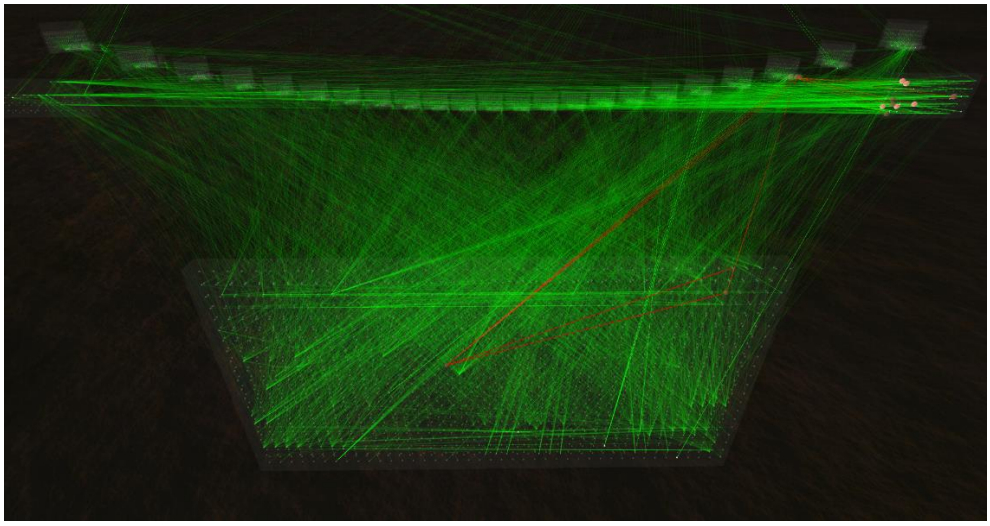

Figure 4: VR view of Locked Shields 18 Partner Run network topology and network traffic using VDE, shown from the other side of the meta-shape, where the data-shape consisting of unknown entities is in foreground (lower side of this screenshot), while Blue Teams' networks are positioned farther away (on the upper side of this screenshot). Some edges and entities have been selected and are rendered red instead of the default green [21].

Two distinct datasets are combined in such an ISPMDV: a logical topology of the entities that are expected to be active in the network (i.e., the positions of nodes representing those entities) and the observed network traffic.

Feedback from analysts on the ISPMDV shown in Figures 2-4 is covered in [27]. Overall, the impressions of stereoscopically perceivable 3D data visualizations were highly favorable, with multiple participants acknowledging that such 3D visualizations of network topology could assist in their understanding of the networks they use daily. Study participants expressed a wish to integrate such visualization capabilities in their workflow. Videos of VR and MR sessions with VDE, as well as some prior conference presentations featuring that tool, are available at coda.ee/JDST.

Figure 5: MR view of Locked Shields 18 Partner Run network topology and network traffic using VDE. A user's index finger is selecting a Blue Team's network [21].

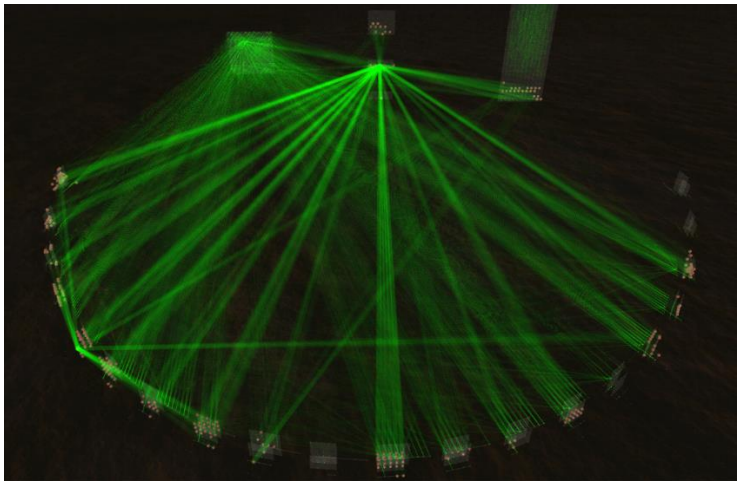

Figure 6: VR view of Locked Shields 16 network topology and traffic using VDE. Notice the slightly different constellation layout compared to Figures 2 - 5 [29].

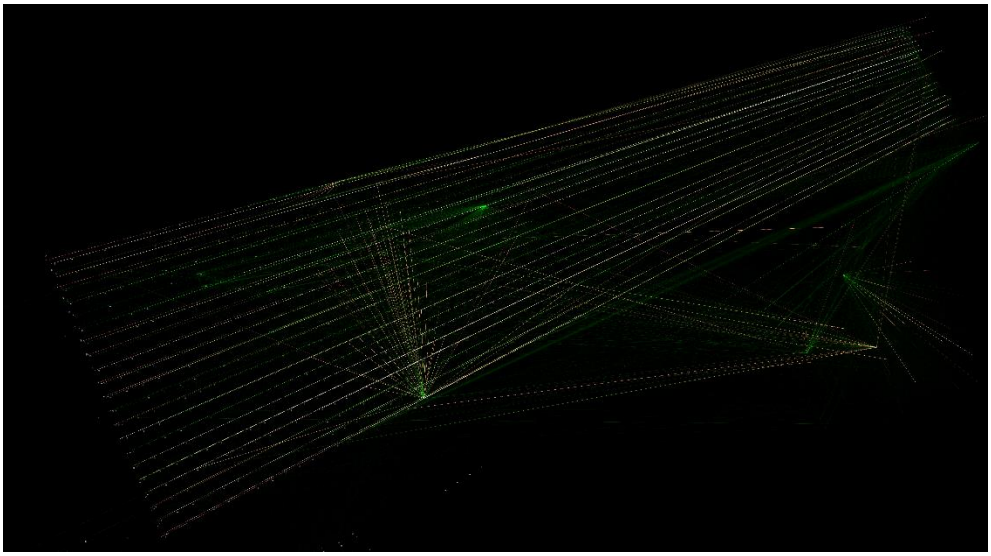

Figure 7: VR view of Locked Shields 16 network topology and traffic using OpenGraphiti. Blue Teams' networks are aligned onto "blades" consisting of subnets, while nodes are positioned on a line sequentially, according to their last octet.

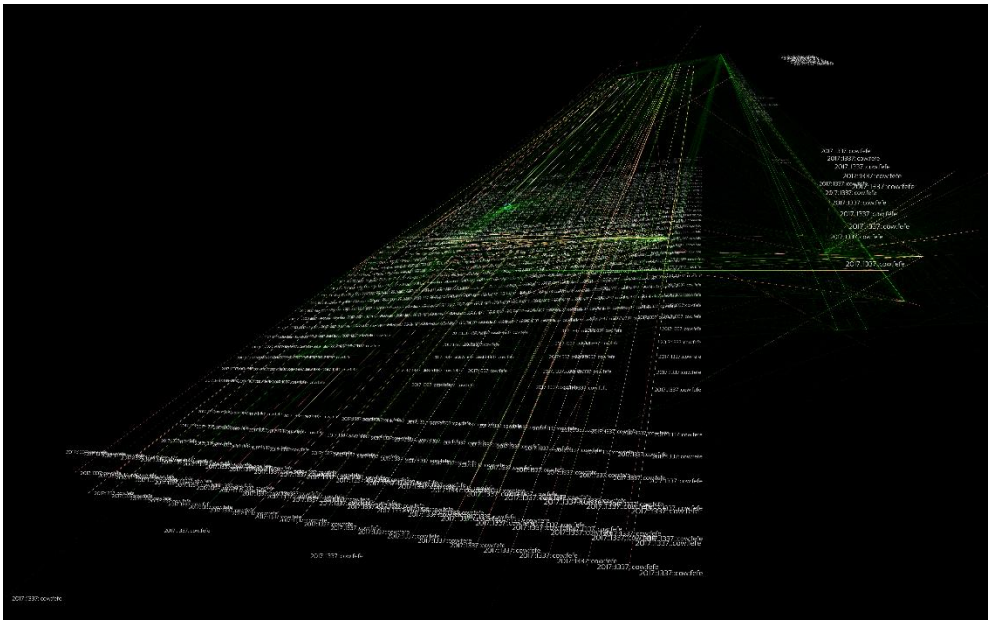

Figure 8: VR view of Locked Shields 16 network topology and traffic using OpenGraphiti. Such layouts are simple to create from network traffic and are useful for initial exploration

of a dataset's topology (after or together with graphs), but are too messy for spotting more subtle anomalies.

B. VDE Demo Dataset

The lack of a public dataset containing the traffic of a computer network with sufficiently complex topology motivated the creation of a mock-up dataset of an imaginary credit union (CU) to showcase a possible network topology ISPMDV, which was then modeled in VDE. This mock-up CU dataset features a financial institution with operations on multiple continents and countries, with multiple branches in each of those, where the branches have standardized, but distinctly populated, internal networks.

Figures 9-14 are screenshots from a VDE v2 VR session, exploring the ISPMDV of the mock-up CU dataset. Video of this exploration was presented at MAVRIC 2020 [25], a VDE v2 demo build is available to experience it in VR [26] and VDE is also included in the NASA MRET open source toolset [6].

In this ISPMDV, subnets of branch networks are grouped to cubes (see internals of that data-shape in Figure 11) which are then stacked vertically based on the organizational group to which that branch belongs (e.g., country, continent). The vertical branch groups are then positioned on a ⅔ circle (Figure 10), with groups containing public services facing the center or the circle. In the center of the ISPMDV are three other groups:

a) known entities (corporate net, partners, etc.),
b) known threats (IP addresses from threat feeds, prior compromises, etc.),
c) unknown IPs.

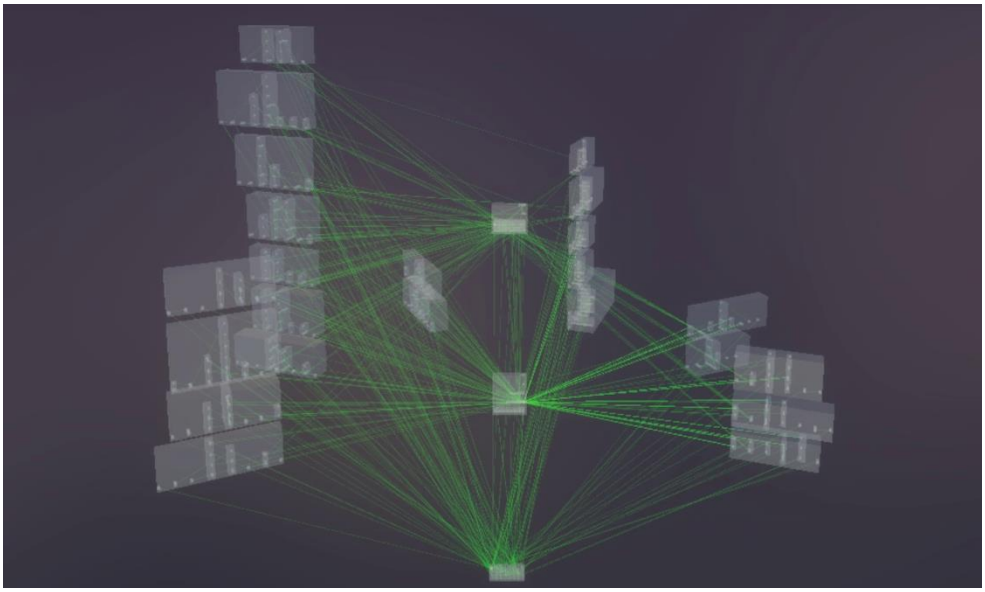

Figure 9: VDE VR sessions of exploring an imaginary CU network's ISPMDV, arranged as a constellation of data-shapes representing the functional topology of that network, overlaid with network traffic.

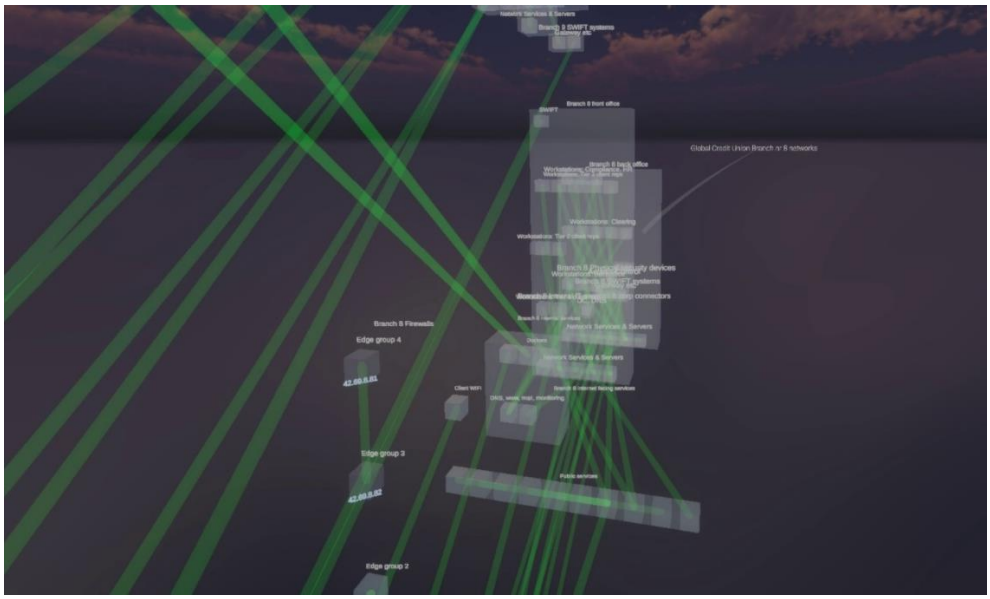

Figure 10: When the user moves the viewpoint closer to one of the data-shapes representing a CU branch network, the outer (cube) shell disappears, while labels of internal groups are activated. Labels of nodes (transparent cubes) are activated only once the user is close enough and are kept facing the user for readability.

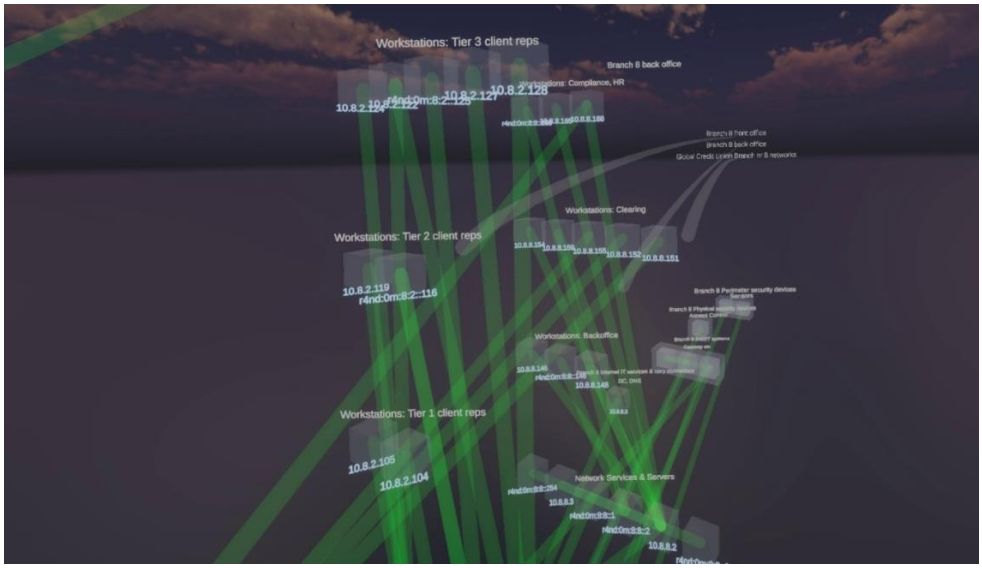

Figure 11: Although subgroup outer shells disappear once the user is close enough (to reduce visual clutter and let the user to focus on individual entities / nodes), subgroups labels (e.g., "Workstations: Backoffice") are kept visible above those, and groups labels (e.g., "Branch 8 front office") are activated based on the direction of the user's gaze.

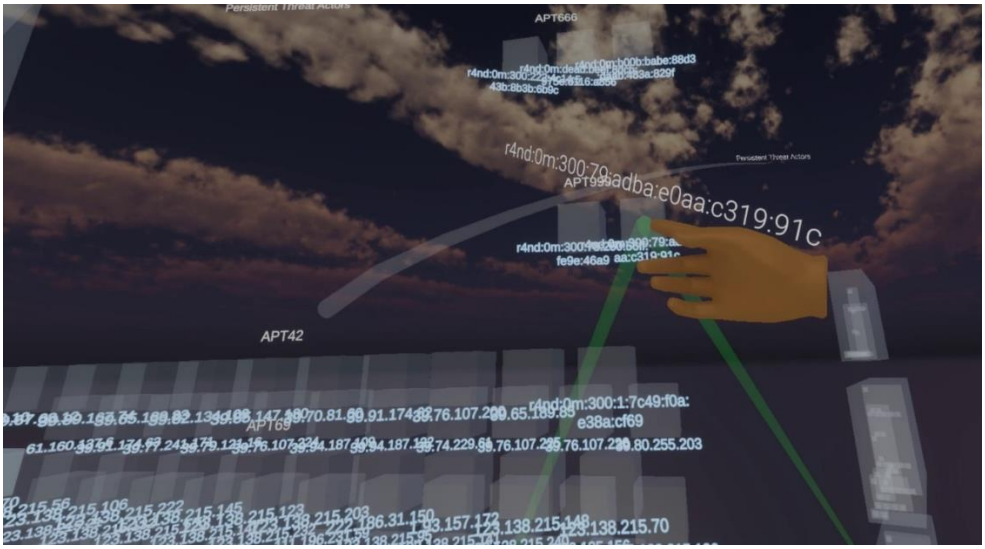

Figure 12: The user can select nodes either (1) from afar, with a pointer or (2) by touching them with their virtual index finger (rendered based on inputs from a VR controller). The selected node's name (in this case, the IP address) is displayed next to the VR hand. The node's incoming and outgoing edges are kept visible while other edges are disabled.

Figure 13: Edges (representing an observed network session) are highlighted when touched, with the corresponding source and destination nodes' names appearing above the user's hand. A single line of text is attached to the hand, avoiding the clutter which a head-up display would have caused. User studies have shown this to be an intuitive feature of the interface.

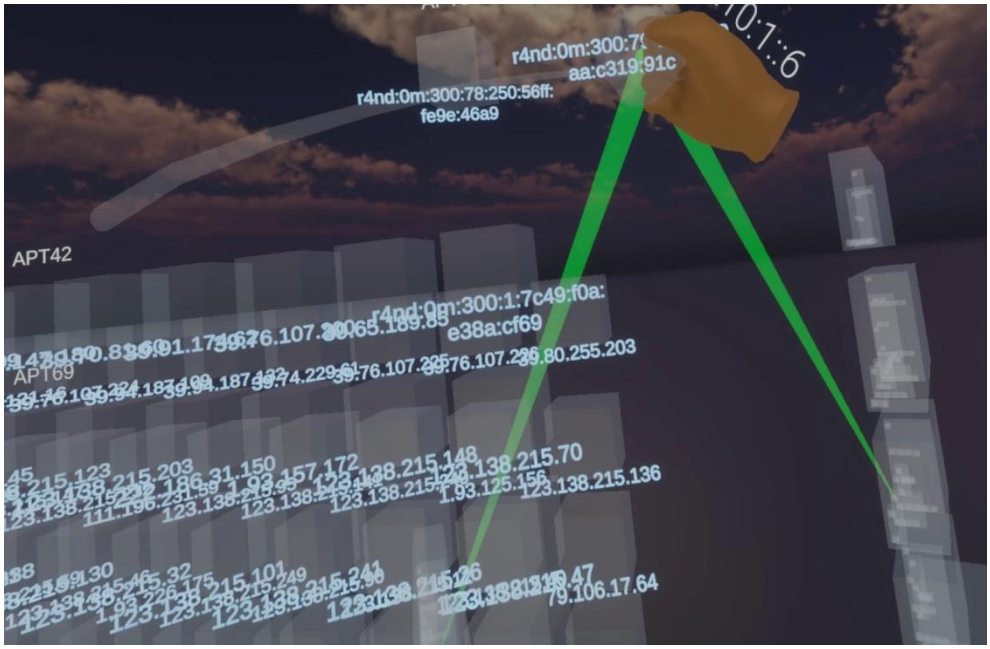

Figure 14: The user can grab a node and move it around, to better perceive the location of the targets and sources of its connections (i.e., terminal points of the edges are easier to spot this way).

The demo build of VDE containing the ISPMDV shown on Figures 9-14 could be used for further studies of user interaction. Together with the VDE server component, VDE can be used to visualize data ingested from SIEM, log correlation, or other data sources' APIs. Please feel free to reach out to the authors to discuss academic research collaboration.

VI. Conclusions

While SPMDVs for intrinsically spatial data have received substantial publicity, the creation, presentation, and usability research of SPMDVs and ISPMDVs designed to show non-spatial data has attracted less attention. In this paper, we explored three distinct ISPMDV examples, all rendered with VDE, with each being used to visualize computer network traffic and topology.

We encourage cybersecurity professionals and researchers to use emerging technologies (e.g., xR HMDs) to explore novel ways for visualizing datasets relevant

to their problems and tasks. The examples provided in this paper are just modest illustrations of what is already possible with existing tools (see [26] [16] [20] [19] [6] and others) and should be used for inspiration.

Appropriate methods (e.g., [8], [24]) should be used when creating ISPMDVs to ensure the utility of the resulting visualization for the CSMEs who would be using them.

ACKNOWLEDGEMENTS

The authors thank Alexander Kott, Jennifer A. Cowley, Lee C. Trossbach, Matthew C. Ryan, Jaan Priisalu, and Olaf Manuel Maennel for their ideas and guidance. This research was partly supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-17-2-0083 and in conjunction with the CCDC Command, Control, Computers, Communications, Cyber, Intelligence, Surveillance, and Reconnaissance (C5ISR) Center.

REFERENCES

[1] Unity Technologies, "Definition of: Virtual Reality (VR)," [Online]. Available: https://unity3d.com/what-is-xr-glossary#paragraph70. [Accessed 2021].

[2] Unity Technologies, "Definition of: Augmented Reality (AR)," [Online]. Available: https://unity3d.com/what-is-xr-glossary#paragraph12. [Accessed 2021].

[3] Unity Technologies, "Definition of: Mixed Reality (MR)," [Online]. Available: https://unity3d.com/what-is-xr-glossary#paragraph42. [Accessed 2021].

[4] S. Skolnik, "Using Virtual Reality to Visualize Disasters, Climate, and Extreme Weather Impacts Shayna Skolnik," in MAVRIC, College Park, 2020.

[5] C. Hurter, Image-Based Visualization: Interactive Multidimensional Data Exploration, N. Elmqvist and D. Ebert, Eds., Morgan & Claypool, 2016.

[6] National Aeronautics and Space Administration, "Collaborative Mixed-Reality Engineering Tool (MRET)," [Online]. Available: https://techport.nasa.gov/view/95677. [Accessed 2021].

[7] A. Kabil, T. Duval and N. Cuppens, "Alert Characterization by Non-expert Users in a Cybersecurity Virtual Environment: A Usability Study," in International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Lecture Notes in Computer Science, 2020.

[8] K. Kullman, L. Buchanan, A. Komlodi and D. Engel, "Mental Model Mapping Method for Cybersecurity," in 22nd International Conference On Human-Computer Interaction, Copenhagen, 2020.

[9] C. Ware and G. Franck, "Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions," ACM Transactions on Graphics, vol. 15, no. 2, pp. 121-140, 4 1996.

[10] J.-P. van Riel and B. Irwin, "InetVis, a Visual Tool for Network Telescope Traffic Analysis," in AFRIGRAPH 2006, Cape Town, 2006.

[11] R. Marty, Applied Security Visualization, 2008.

[12] T. Munzner, Visualization Analysis & Design, A K Peters/CRC Press, 2014, p. 428.

[13] H. S. Smallman, M. St. John, H. M. Oonk and M. B. Cowen, "Information availability in 2D and 3D displays," IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 51-57, 2001.

[14] M. Teräs and S. Raghunathan, "Big Data Visualisation in Immersive Virtual Reality Environments: Embodied Phenomenological Perspectives to Interaction," ICTACT Journal on Soft Computing, vol. 05, no. 04, pp. 1009-1015, 2015.

[15] A. Kabil, T. Duval, N. Cuppens, G. L. Comte, Y. Halgand and C. Ponchel, "Why should we use 3D Collaborative Virtual Environments for Cyber Security?," in IEEE Fourth VR International Workshop on Collaborative Virtual Environments, Reutlingen, 2018.

[16] M. Cordeil, A. Cunningham, B. Bach, C. Hurter, B. H. Thomas, K. Marriott and T. Dwyer, "IATK: An Immersive Analytics Toolkit," in IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, 2019.

[17] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas and K. Marriott, "There Is No Spoon: Evaluating Performance, Space Use, and Presence with Expert Domain Users in Immersive Analytics," IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 536 - 546, 2020.

[18] S. Beitzel, J. Dykstra, P. Toliver and J. Youzwak, "Exploring 3D Cybersecurity Visualization with the Microsoft HoloLens," in International Conference on Applied Human Factors and Ergonomics, 2017, 2017.

[19] T. Reuille, S. Hawthorne, A. Hay, S. Matsusaki and C. Ye, "OpenDNS Data Visualization Framework," 2015. [Online]. Available: http://www.opengraphiti.com/.

[20] 3Data, "Advanced Analytics for SecOps," 3Data, [Online]. Available: https://3data.io/solutions-cybersecurity/. [Accessed 01 2021].

[21] M. Ryan, K. Kullman and L. Trossbach, "VR/MR Supporting the Future of Defensive Cyber Operations," in NATO Computer Aided Analysis, Exercise, Experimentation Forum., Paris, 2019.

[22] Y. Seong, J. Nuamah and S. Yi, "Guidelines for Cybersecurity Visualization Design," in IDEAS 2020, Seoul, South Korea, 2020.

[23] C. Zhong, A. Alnusair, B. Sayger, A. Troxell and J. Yao, "AOH-Map: A Mind Mapping System for Supporting Collaborative Cyber Security Analysis," in 2019 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), Las Vegas, 2019.

[24] D. J. Clark and B. P. Turnbull, "Experiment Design for Complex Immersive Visualisation," in Conference: Military Communications and Information Systems Conference (MilCIS) 2020, Canberra, 2020.

[25] K. Kullman, "Creating Useful 3D Data Visualizations for Cybersecurity," in MAVRIC, College Park, MD, 2020.

[26] K. Kullman, "Virtual Data Explorer," Cognitive Data OÜ, [Online]. Available: https://coda.ee/getvde.

[27] K. Kullman, N. B. Asher and C. Sample, "Operator Impressions of 3D Visualizations for Cybersecurity Analysts," in ECCWS 2019 18th European Conference on Cyber Warfare and Security, Coimbra, 2019.

[28] The NATO Cooperative Cyber Defence Centre of Excellence, "Locked Shields Cyber Defence eXercise," [Online]. Available: https://ccdcoe.org/exercises/locked-shields/.

[29] K. Kullman, J. Cowley and N. Ben-Asher, "Enhancing Cyber Defense Situational Awareness Using 3D Visualizations," in 13th International Conference on Cyber Warfare and Security, Washington, DC, 2018.