BOM-Vis: A Visualization of Network Health and Status

Dennis Park

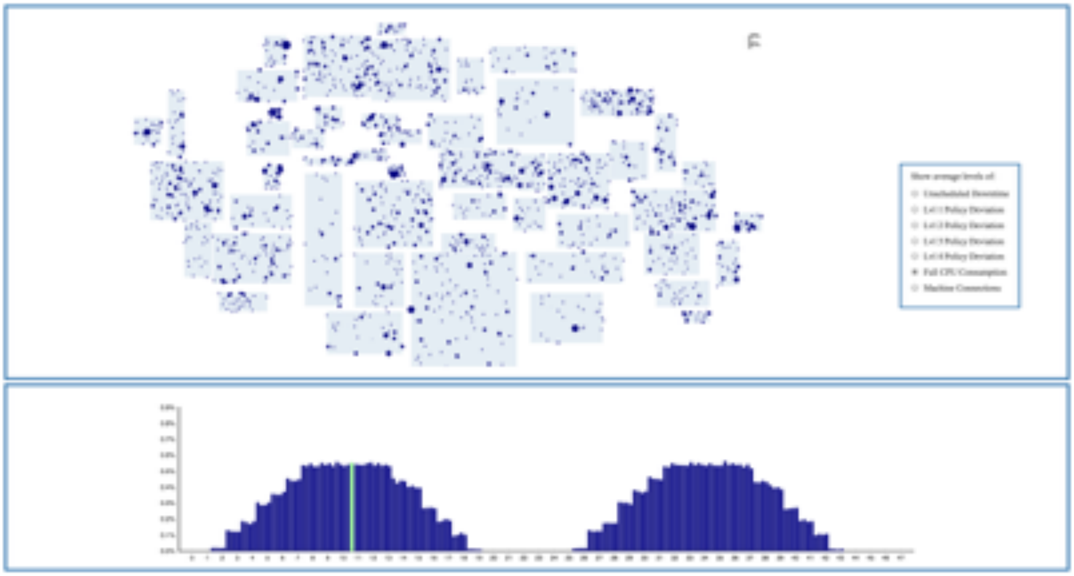

Fig 1. The main screen showing together the Static View (top) and Dynamic View (bottom). Through the Static View, users can view the statistics of the network at a single point in time laid out over a geographical representation of the network's various regions. Through the Dynamic View, users can view the statistics of the network as it evolves over time. By toggling a set of radio buttons (right), the user is able to choose which aspect of the network to visualize.

Abstract—As businesses and organizations expand, their networks evolve correspondingly, growing both in their size and their complexity. Coupled with the increasing threats to network security and the increasing costs of network failures, administrators require sophisticated tools for monitoring the health and status of their networks which can meet the demands of scalability. In this paper, I present BOM-Vis, a visualization tool for monitoring network health and status. Using data provided by the 2012 VAST challenge, BOM-Vis serves as a case study in how scalable visualizations of network health and status can be designed.

1 Introduction

As computer networks grow in size and complexity, it becomes increasingly difficult to effectively convey the gathered monitoring data to human users. Coupled with increasing security threats and increasing costs of failure, network administrators require tools which can effectively support the consumption of such large volumes of network information.

Information visualizations have been put forward as an essential tool in meeting these challenges of network monitoring [1]. Accordingly, the visualization of computer networks are a rich area of research within the InfoVis community [1].

In the current paper, I consider the problems of the 2012 VAST Challenge as a useful instance of computer network visualization problems with their difficulties of scale.

The Visual Analytics Science and Technology (VAST) Challenges are annually held competitions which provide researchers and software teams with fictional but realistic visualization tasks and datasets to test their visualization software, and encourage innovation in solving visualization problems of greater and greater complexity [2].

In the 2012 VAST Challenge, we are introduced to BankWorld, a planet much like Earth, identical in size, but with a geography consisting of a single land mass in which a handful of nation-states exist side by side. Within BankWorld, the Bank of Money (BOM) is the most important organization, having its facilities spread out all across the globe. BOM is organized into a collection of regions: large regions, small regions, and HQ, each populated by various facilities such as data centres (overseen by HQ), regional headquarters, and branches of the bank.

The BOM network, which is the focus of this challenge, is made up of various machines housed all across the facilities of the organization. These machines are first categorized into different classes, such as "servers," "workstations," and "atms." They are then further categorized in terms of functionality. For example, servers can be any one of "web servers," "email servers," "file servers," "compute servers," or "multiple functionality servers."

The machines of this network are queried for status reports during every 15 minute intervals, describing various aspects of the machines' status and health such as the number of connections they currently have open, whether they are experiencing full consumption of cpu, etc.

The subject of focus for the 2012 VAST Challenge is a 48 hour interval just prior to a network wide failure. In this paper, I present BOM-Vis, a tool for visualizing the health and status of a network, as a solution for investigating and understanding the causes of the network failure by supporting the discovery and profiling of anomalies in the network, both over time and across region.

2 Related Work

As a past VAST Challenge, there already exists a wealth of different solutions which have been submitted by various teams around the world. In the following, I discuss various design decisions made in these past solutions, and discuss their relation to the decisions made in designing BOM-Vis. In particular, I consider how the various solutions have gone about solving two key problems involved in designing a network monitoring visualization: (i) *How do you deal with the temporal aspect of a network's health and status; in other words, how is the changes in the network over time visualized?* (ii) *How do you present the numerous aspects of a network's health and status to the user in a way that does not overload the user with information?*

2.1 Visualizing Temporal Changes

In visualizing changes in a network over time, past solutions have generally taken one of two approaches: (i) provide a timeline showing the value of a network attribute as a function of time or (ii) provide an animation which shows a frame by frame evolution of the network over time. Solutions which have gone the first route [3-5], have the advantage of showing the evolution of multiple network attributes at once, allowing the user to discover correlations between attributes. In contrast, solutions going the second route such as that submitted by the BusinessForensics team [7], allows the user to view how individual regions evolve over time, rather than having their data be aggregated into network-wide summaries and consequently lose much of the detail.

At the same time, these solutions suffer from their inability to provide a clear overview of the network's change over time. Some solutions, such as that submitted by the Secure Decisions team [3], counterbalance the loss of information due to aggregation by allowing users to select a particular point in a network-wide timeline in which to drill down and break the aggregated attribute value down to its component parts (e.g. the number of machine connections per business unit). In this way, by providing the user with an overview of how the network changes over time, then allowing them to drill down into a particular time-point for detailed information, the disadvantages of the two aforementioned routes are mitigated.

In designing BOM-Vis, I borrow from the Secure Decisions solution and provide an overview of the network's change over time in which users can select time-points to drill down into. Unlike the Secure Decisions solution, however, I provide a breakdown of an attribute into regional contributions. In this way, the system is able to support a back and forth "drilling down" between the static and dynamic views, whereby a user selects a particular time-point to show in greater static detail, then using this greater level of detail select a particular region for which to show dynamic information in isolation from the rest of the network.

2.2 Dealing with Numerous Attributes

In providing a snapshot of the network's attributes at a given point in time, most solutions have tended to encode the values of machine attributes onto a geographic map of the network so as to match the user's mental model of the network as it exists in the world. The variability in solutions have mainly come in the details of how these attributes are encoded on the map. Several solutions provide a panel of attributes which they can turn on or off [4], which then maps the various selected attributes simultaneously onto the static view using separate channels of encoding. Others have instead reserved the static view solely for a single attribute, such as changes in policy levels, and show other attributes by other, time-dependent views [3].

The current solution has favoured the former approach, although with the difference that rather than providing a panel of checkboxes to view multiple attributes simultaneously, it instead provides a set of radio buttons used to toggle between the various attributes. The obvious disadvantage with this approach is that finding correlations between different attributes become difficult, requiring users to constantly toggle back and forth between the attributes of interest. However, because the current system is less interested in finding relationships between attributes, and more interested in finding points of anomaly, the radio buttons were chosen for their added simplicity. Because, only a single attribute is ever shown on the page, the system is able to encode them uniformly, not requiring its users to learn the semantics of multiple encodings.

3 Tasks

The following provides a detailed listing of the core tasks I intend to support:

- Identify regions experiencing significant levels of unscheduled downtime at a given moment in time.
- Having identified a region experiencing significant levels of unscheduled downtime, identify when it began and when it ended (i.e. identify the interval of anomaly).

- Identify regions experiencing significant levels of policy deviation at a given moment in time.
- Having identified a region experiencing significant levels of policy deviation, identify when it began and when it ended (i.e. identify the interval of anomaly).
- Having identified an interval of anomaly w/r/t policy deviation, profile its growth (e.g. where it began, how

it spread from facility to facility and region to region).

- Identify regions experiencing significant levels of full cpu consumption (relative to other regions) at a given moment in time.
- Having identified a region experiencing significant levels of full cpu consumption (relative to other regions), determine whether the levels are anomalous for the given region by viewing it within the context of time (i.e. does it deviate from the region's normal levels of cpu consumption?).
- Having confirmed that a region experienced anomalous levels of full cpu consumption, identify when it began and when it ended (i.e. identify the interval of anomaly).

- Identify regions experiencing anomalous levels of machine connections (relative to other regions) at a given moment in time.
- Having identified a region experiencing anomalous levels of machine connections (relative to other regions), determine whether the levels are anomalous for the given region by viewing it within the context of time (i.e. is the time-point of focus a spike or dip breaking the general pattern of connection levels in the given region?).
- Having confirmed that a region experienced anomalous levels of machine connections, identify when it began and when it ended (i.e. identify the interval of anomaly).

- Identify an interval of anomaly w/r/t machine connection levels, which breaks the general pattern (i.e. identify spikes and dips breaking the general pattern of connection levels). Similarly, identify an interval of anomaly w/r/t unscheduled downtime, policy deviation, and full cpu consumption levels.
- Having identified an interval of anomaly w/r/t machine connection levels, localize the anomaly to a particular region or group of regions. Similarly, having identified an interval of anomaly w/r/t/ unscheduled downtime, policy deviation, or full cpu consumption levels, localize the anomaly to a particular region or group of regions.

4 Data

4.1 Data Acquisition

As previously mentioned, the dataset for BOM's network is provided by the IEEE Conference of Visual Analytics Science and Technology (VAST) as part of their 2012 visualization challenge.

Two separate datasets are provided: (i) a catalogue of BOM's 895,025 machines, detailing their locations, IP addresses, types, and functions; (ii) and a log of status reports spanning a total of 48 hours using 15 minute intervals, which describe the health, activity levels, and status of the machines across the network.

4.2 Data Reduction

One major problem with the original dataset as provided by the challenge is its enormous size. With 192 distinct time-points, each with a status report for most of the 895,025 machines, the dataset reaches close to 2 GB of data. The first decision needed to be made in developing the current solution was to reduce this original dataset into something more manageable.

In order to support users in gaining an overview of the dataset and then drilling up and down levels of details as needed, the original log of machine status reports was transformed into 3 separate tables of data, corresponding to the 3 main levels of abstraction: (i) a log of facility-wide status reports (ii) a log of region-wide status reports (iii) a log of network-wide status reports.

Because I wanted to display the statistics of a given time-point over a geographic representation of BankWorld's regions, I needed to derive coordinates for these regions as well as for the coordinates for the facilities. These coordinates were derived by first noticing that the machines belonging to a single facility shared the same coordinate. By grouping them according to their facilities, the coordinate of these facilities were easily derived, and then regional coordinates were created by taking the leftmost,rightmost,topmost, and bottommost facilities of each region and representing them as rectangular.

Another derivation made was regarding the statistics for machine downtime. As mentioned in the challenge description, when a status report for a machine is missing for a particular time-point, this indicates that the machine was experiencing downtime at the moment and was unresponsive to the queries for a status report. This fact was used to derived statistics for downtime by identifying the number of missing status reports for each time-point.

In this way, the post-processed dataset can be understood as a time-based geometric dataset consisting of various points on a two-dimensional grid (i.e. the planet) with a collection of associated values (i.e. machine attributes). The dataset can be described in further detail by defining over the points a partitioning into the set of regions in which the facilities exist.

In the current solution, the temporal aspect of the dataset and the spatial aspect of the dataset are treated separately by having a pair of views which are shown together on the screen.

5 Solution

As previously mentioned, the purpose of this project is to design and implement a visualization tool, which can be used to monitor the health and status of BOM's network, and in particular discover and understand the anomalies in the network which arise. The current solution tries to achieve this by a combination of two main views—one providing a dynamic picture of how the network changes over time, and another providing a static picture of the network at a single moment in time—presented one below the other. These views are described next.

5.1 Static View

The Static View provides an overview of the network's health and status at a given moment in time. It consists of a BankWorld map partitioned into its various regions populated by small points to represent facilities, along with a set of radio buttons which can be toggled to focus on different aspects of the network's status.

When the option for *unplanned downtime* is selected, the various facilities of BOM become encoded by the size of their representative marks to show the percentage of machines housed in each facility that are experiencing downtime.

When the option for *lvl 1 policy deviation* is selected, the facilities again become encoded by mark size to represent the percentage of machines experiencing level 1 policy deviations. The options for *lvl 2 policy deviation*, *lvl 3 policy deviation*, and *lvl 4 policy deviation* are implemented similarly.

When the option for *full cpu consumption* is selected, the facilities become encoded by mark size again to represent the percentage of machines experiencing full cpu consumption.

When the option for *machine connections* is selected, the facilities become encoded by mark size to represent the mean connection level of the machines housed in each facility.

Beyond the features described above, the static view also interacts with the *dynamic view* (described below) to navigate along different points in time. By selecting different columns on the *dynamic view*, the user is able to jump between different points in time and see the statistics for the corresponding time-point drawn on the Static View. The currently selected time-point will be represented by a highlighted bar in the Dynamic View

5.2 Dynamic View

This view provides a representation of the network's health and status across time. It consists of a horizontal timeline representing the 48 hours of activity logged in the provided dataset. As with the static view described above, the dynamic view changes in response to user toggling of the radio buttons to show different aspects of the network.

The dynamic view is also dependent on user interaction with the static view in the following way: when users click on a region within the static view, it becomes the focus of the dynamic view, showing time-dependent data for just that region. In this way, the Static View serves as a panel of controls for filtering the data shown on the Dynamic View. When a selected region is unselected, the Dynamic View returns back to showing the aggregated network-wide data. Initially, all regions are unselected, and the dynamic view shows time-dependent data aggregating the entire network.

When the option for unplanned downtime is selected, the dynamic view shows a bar chart encoding the percentage of machines experiencing unplanned downtime in the network/region as a function of time.

When the option for level 1 policy deviation is selected, the dynamic view shows a bar chart encoding the percentage of machines experiencing level 1 policy deviations in the network/region. Level 2-4 policy deviations are implemented similarly.

When the option for full cpu consumption is selected, the dynamic view shows a bar chart encoding the percentage of machines experiencing full cpu consumption in the network/region.

When the option for machine connections is selected, the dynamic view shows a line chart encoding the average machine connection levels for the network/region as a function of time.

6 Implementation

To create BOM-Vis, I used the following tools and libraries:

- **D3.js: Data-driven Documents.** This library was used to create both the Static View and Dynamic View.

- **jQuery.** This library was used along with d3 to deal with user interactions, such as the radio button toggling, temporal navigation, and regional filtering.

- **Bootstrap.** This framework was used to layout the views on the page.

- **MySQL.** This was originally intended to house the data and be retrieved from the application, however once the data was reduced to a manageable size, the data was instead exported as cvs files and kept within the server directory. MySQL was however used to transform the data.

- **Node.js** This was used as a basis of the backend server.

- **Express.js** This was used to build a simple static server.

7 Scenario Walkthrough

On February 2, 2012, the Bank of Money network experienced network wide outage. John has been brought in to find out how it may have happened.

He starts by running BOM-Vis. Once it loads, he moves to the very last recorded status report by clicking the far right end of the dynamic view. He then toggles through the various options (*unplanned downtime*, *lvl 1-4 policy deviation*, *full cpu consumption*, *machine connections*), getting an understanding of the network's state at this last time-point. He toggles back through the *policy deviation* options. In this last time-point, the network seems to have experienced high levels of policy deviations (for all levels) all across the network. By looking through the dynamic view, John tries to identify when the policy deviations began.

Each of the policy deviation levels seem to have grown linearly through the 48 hours, with the onset of the higher levels of deviation coming later.

The one exception to this is when John toggles back to the level 1 policy deviation option, which shows that the level 1 policy deviations was present from the very beginning of the 48 hours.

Furthermore, he sees on the static view that there seems to be two regions in particular, region 5 and 10, which seems to be exhibiting level policy deviations at much more extreme levels that the other regions. To investigate, he selects region 5 on the static view, thereby filtering the dynamic view to show only the data for this region. When he does so, he sees that level one policy deviations started at maximum levels for region 5 at the very beginning of the 48 hours. He selects the first time-point on the dynamic view. At this point, the static view shows that only two regions are experiencing level one policy deviations but that they are experiencing at severe levels.

John decides that these two regions are the best suspects for the outbreak of the network problems and goes to contact personnel at these locations to investigate further.

8 Discussion and Future Work

To reiterate the goals of the project, the purpose of BOM-Vis was to take a design study in creating a visualization for network health and status monitoring. In particular, I wanted to develop a solution which would support the discovery and profiling of anomalies in the network.

As illustrated by the previous scenario, I believe that BOM-Vis is capable of supporting the tasks which I set out to support.

There are however several limitations of the current implementation of BOM-Vis which it would be interesting to look into for future work. For instance, the current implementation works with a very small subset of machine attributes from what existed in the original dataset. Many attributes such as the recognition of login failure at certain machines may be interesting to incorporate into the solution, providing a wider array of information for users to make sense of anomalies.

Also, currently BOM-Vis is only able to filter the dynamic view by single regions. It would be interesting to extend this feature so that multiple regions in the static view can be selected and the dynamic view would show the aggregate of those selected regions. Similarly, an additional feature which would be useful would be to allow the selection of intervals of time, rather than single instances of times, so that by dragging an interval over the dynamic view, the static view would display statistics for the regions which aggregate the statistics within the given interval.

9 Conclusion

In BOM-Vis, I created a tool for visualizing the health and status of a computer network, with the focus of discovering anomalies both over time and across regions. It consists of two major views: the Static View shows the statistics of a single point in time laid over a geographic representation of BankWorld's regions, and the Dynamic View shows how attributes of the network change and evolve over time. A set of radio buttons are provided to toggle between different aspects of the network, and by interacting with the static and dynamic view elements, the user is able to filter the network information by both time and region.

References

[1] J. Goodall, F. Mansmann, and J. Gerth. Computer Network Visualization. *IEEE Network,* 2012.

[2] vacommunity.org/VAST+Challenge+2012, 2012. Retrieved December 16, 2015m from http://vacommunity.org/VAST+Challenge+2012.

[3] M. Farry, R. Stark, A. Wollocko, and M. Borys, "CRA-Farry-MC1," *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/Charles%20River%20Analytics/

[4] C. Horn, C. Ellsworth, and D. Halperin, "SecureDecisions-Horn-MC1," *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/Secure%20Decisions/

[5] L. Laberge et al., "GDC4s-Laberge-MC1," *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/General%20Dynamics%20C4%20Systems/

[6] R. Pabst, "BF-Pabst-MC1," *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/BusinessForensics/

[7] J. Gbls-Szab et al., "SZTAKI-DMS: OWLAP Analytics Beta," *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/MTA%20SZTAKI/

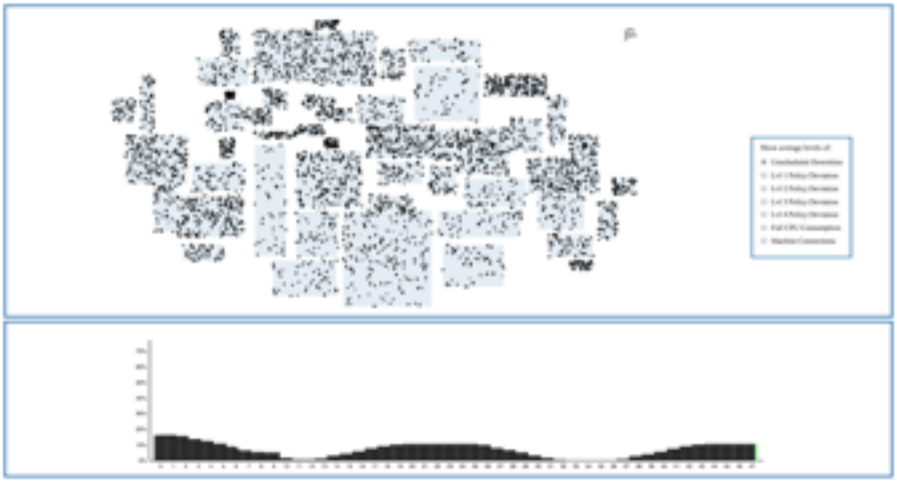

Fig 2. Downtime during the last time-point.

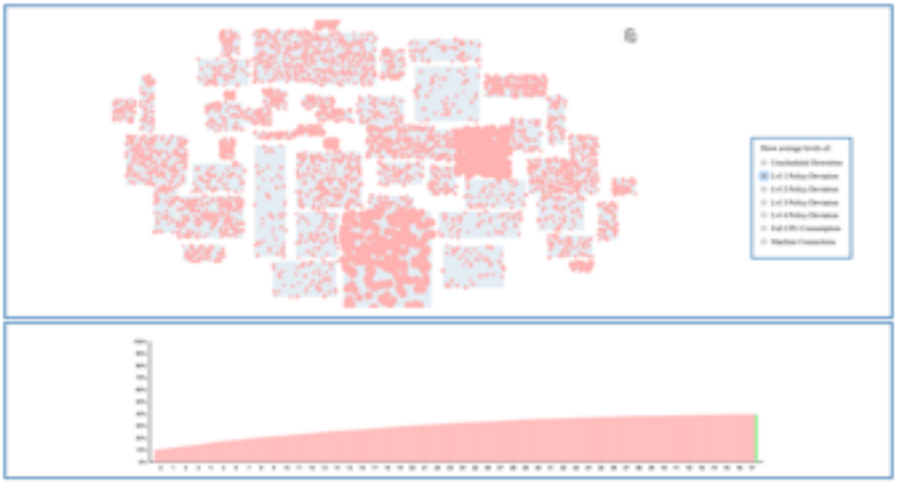

Fig 3. Showing anomalous regions for lvl 1 policy deviation during the last time-point.

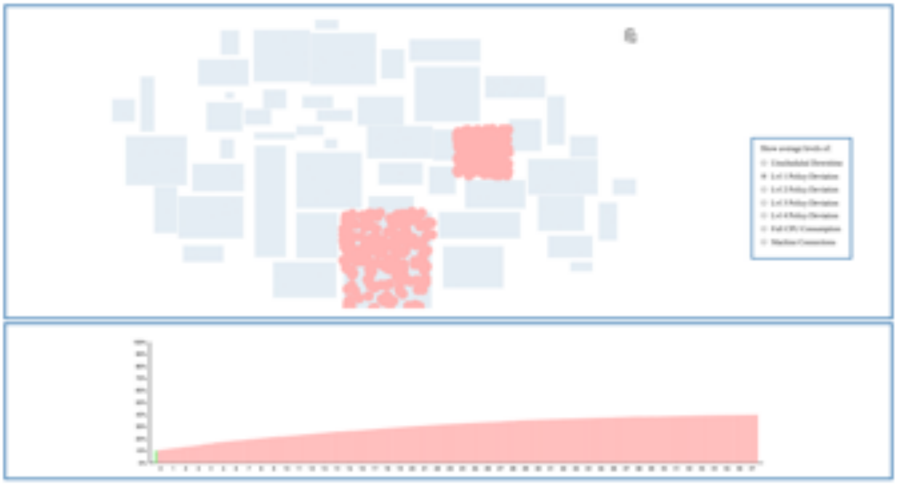

Fig 4. Showing anomalous regions for lvl 1 policy deviation during the first time-point.