Solving the Traffic and Flitter Challenges with Tulip

Paolo Simonetto, Pierre-Yves Koenig, Faraz Zaidi, Daniel Archambault, Frédéric Gilbert, Trung Tien Phan Quang, Morgan Mathiaut, Antoine Lambert, Jonathan Dubois, Ronan Sicre, et al.

Solving the Traffic and Flitter Challenges with Tulip

Award Traffic Mini Challenge: Innovative Visualization and Excellent Description

Award Flitter Mini Challenge: Representation of Uncertainty in Rules & in Visualization

Paolo Simonetto* Pierre-Yves Koenig* Faraz Zaidi* Daniel Archambault† Frédéric Gilbert*
Trung-Tien Phan-Quang* Morgan Mathiaut* Antoine Lambert* Jonathan Dubois*
Ronan Sicre* Mathieu Brulin* Remy Vieux*
Guy Melançon†
INRIA Bordeaux Sud-Ouest and LaBRI, Université de Bordeaux I

Abstract

We present our visualization systems and findings for the Badge and Network Traffic as well as the Social Network and Geospatial challenges of the 2009 VAST contest. The summary starts by presenting an overview of our time series encoding of badge information and network traffic. Our findings suggest that employee 30 may be of interest. In the second part of the paper, we describe our system for finding subgraphs in the social network subject to degree constraints. Subsequently, we present our most likely candidate network which is similar to scenario B.

Index Terms: H.5.0 [Information Systems]: Information Interfaces and Presentation—General; F.2.2 [Theory of Computation]: Nonnumerical Algorithms and Problems—Pattern matching;

1 Introduction

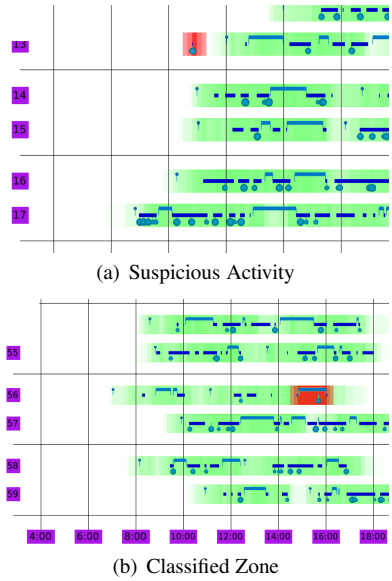
In order to determine the most likely candidates for the embassy leak and the pattern in the social network, we developed two visualization systems using the Tulip [1] graph drawing libraries and software. The first system, presented in section 2, describes our time series visualization to understand the situation at the embassy using a time series of widgets. Section 3 describes our system for finding patterns in the larger social network.

2 Badge and Network Traffic Challenge

Our visualization technique, shown in Figure 1, is based on a timeline view. The horizontal axis encodes the time of the day at hour intervals, while the vertical axis encodes the employee ID and IP address. The horizontal lines in the grid group employees into offices. For example, in Figure 1(a), employees 14 and 15 are in the same office because they are in between the same horizontal lines.

The timeline of each employee collects four kinds of data. First, the upwardly directed glyphs, the teal circles and bars, encode the door log events. Circles are badge-in events. Bars between two vertical lines encode the time interval between a badge-in-classified event and badge-out-classified event for an employee. The central blue bars show when the employee's computer is active. Downwardly directed circles represent transmissions and the size of the circle is proportional to transmission size.

A green background shows the average daily activity of an employee over the thirty-one days. A more saturated green indicates a higher probability that the employee is at work. The colour of this green background can be changed to red, highlighting suspicious activities. The most suspicious activities occur when an employees computer is active but he or she is most likely not at their desk.



(a) Suspicious Activity

(b) Classified Zone

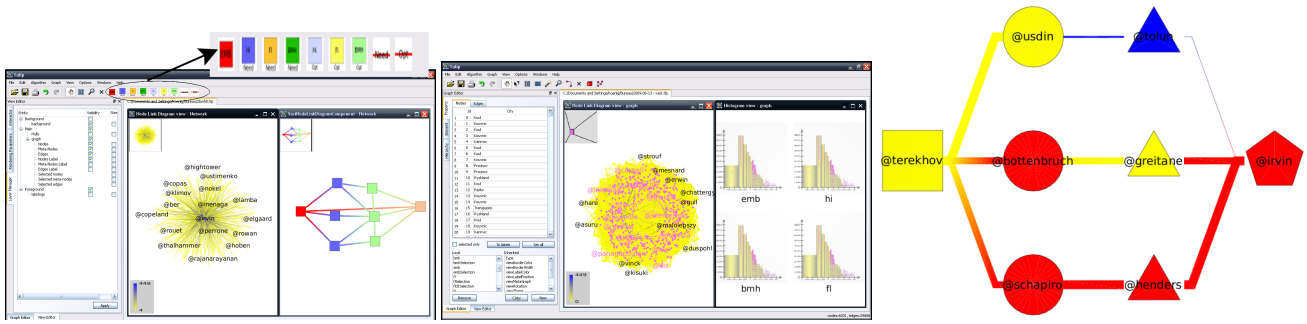
Figure 1: Two suspicious transmissions. **(a)** Employee 13's computer is used for a large upload on day twenty-two before the employee usually arrives at work. **(b)** A case where an employee's computer is used while the employee is in the classified zone.

We found three exceptionally suspicious transmissions in the data. In our first case, Figure 1(a) shows a large transmission from employee 13's computer on day twenty-two, forty minutes before the employee usually arrives at work. The employee had yet to badge into the building and the badge-in-building event recorded was typical for this employee. In our second case, a large transmission is sent less than a minute before employee 20 badges into the building on day twenty-nine, indicating the employee is probably not at their desk. Finally, a large transmission is sent from employee 18's computer over two hours since it was last active on day seventeen. In all three cases, the office was most likely empty as the computer activity and the green background in the visualization indicate. A leak could have been sent from this computer by someone who is not assigned to this office.

Subsequently, we looked for transmissions sent from an employee's computer when the employee was in the classified zone. During this time, the employee's computer should not be used, because they are away from their desk. Figure 1(b) shows an example, but, in reality, we found eight such cases.

*{paolo.simonetto, pierre-yves.koenig, faraz.zaidi, frederic.gilbert, phanquan, mathiaut, antoine.lambert, dubois, sicre, mathieu.brulin, vieux }@labri.fr

†{daniel.archambault, guy.melancon}@inria.fr



(a) Query Interface (b) Degree Selection (c) Candidate Solution

Figure 2: Task 2 visualization system. **(a)** Visual Query Generator interface with the tool bar showing Employee(EMB), Handlers(Hi), Middlemen(BMH), Fearless Leader(FL) nodes and required or optional edges. **(b)** Histogram view allowing user to choose the range for the number of connections for each role. **(c)** Our best candidate solution for this task.

All eleven of these suspicious transmissions were sent to the same IP address, `100.59.151.133`, and on the same port, 8080. We figured that this machine is where the leaks were uploaded. Subsequently, we highlighted all transmissions to this IP address and found seven other transmissions. The office was probably empty in all cases with one interesting exception: employee 30 was most likely in 30/31's office when three of the suspicious transmissions were sent from employee 31's computer. After plotting all suspicious transmissions on the embassy map, we found that they were clustered around office 15: employee 30's office.

Finally, we found five cases where an employee either badged into the secure zone without badging out or vice versa, and employee 30 was involved in three of these five infractions.

3 Social Network and Geospatial Challenge

The interactive query generator, shown in figure 2(a), enables the user to generate the pattern to be found in the larger graph through a visual interface. Using the supplied input, the system applies constraints to the graph, simplifying the search for the given pattern.

The tool bar contains nodes with different roles and edges, allowing the user to construct patterns. The user can choose any number of handlers, middlemen, and fearless leaders. These nodes are connected through optional or required edges. Required edges play an important role in filtering as they impose minimum degree constraints for each role. The Employee (EMB), handlers (Hi), middlemen (BMH) and fearless leader (FL) all have different degree constraints. Through a histogram view, shown in figure 2(b), the user chooses degree ranges for each role based on the degree distribution in the entire graph.

For example, in the task description, the degree of an employee is about 40. In the histogram view, few nodes have a degree above 45. Therefore, we take 45 as an upper bound for employee degree and the same level of uncertainty as a lower bound. Handler candidates have a degree range between 30 and 40. We enter the range [27, 43], ensuring that no solution near these bounds would be missed. We estimate that the degree of a middleman is about 6. These middlemen communicate with one or more handlers and there are 3 at most. Middlemen have one or two more contacts, one of them being the fearless leader. Therefore, a middleman has a degree of about 5 (3 handlers + 1 or 2 other contacts). Thus, we set a maximum degree of 6, which allows for some uncertainty. Finally, the fearless leader has over 100 contacts which we use as a lower bound.

To search for the patterns in the graph, the algorithm creates sets of possible candidates for each role based on the degree constraints and iteratively removes elements from these sets based on connectivity constraints. The induced subgraph of the union of these sets is computed to filter the network. Additional connectivity constraints are applied from the given pattern drawn in the query generator. For example, an employee must be connected to three handlers, thus, any node having a degree less than three is removed. Similar constraints can be applied to handlers, middlemen, and the fearless leader.

Connection constraints exist between all roles: the employee is connected to the handlers, the handlers to the middlemen, and the middlemen to the fearless leader. We use these constraints to filter the network further. Once complete, our system constructs separate graphs for each possible candidate network.

From this point forward, we manually filter the candidate networks. The system uses shape to encode each role: squares for employees, circles for handlers, triangles for middlemen, and pentagons for fearless leaders. The constraint that allowed us to eliminate the most candidate networks was that handlers had to be directly connected to the employee and must communicate with the fearless leader through a middleman. In any of the candidate networks, if we see a fearless leader doesn't communicate with exactly three handlers then the candidate network is rejected.

To filter this list of solutions further, we use the geospatial information. Nodes in the candidate network are coloured by city size according to the legend on the Flovania map. Distances between cities are mapped to edge width where thicker edges correspond to nodes that are close geospatially. Our best candidate solution to the Flitter challenge is shown in figure 2(c). As the leader is in a larger city, the fearless leader must be coloured either yellow or red. Middlemen are in a city geospatially close to the employee or to the handler, and, thus, should have thick edges in these connections.

4 Conclusion

We believe that employee 30 probably caused the embassy leak, because most of the suspicious transmissions were sent from locations near the employee's office, and he or she is one of the few employees who could have sent all of these transmissions.

We, also, believe that employee working at the embassy has a Flitter name @terekhov. Our candidate solution matches scenario B where each handler has his or her own middleman. The fearless leader in this network has Flitter name @irvin.

Acknowledgements

We thank the INRIA the Gravité project for supporting this work.

References

[1] D. Auber. Tulip : A huge graph visualization framework. In P. Mutzel and M. Jünger, editors, *Graph Drawing Software*, Mathematics and Visualization, pages 105–126. Springer-Verlag, 2003.