Badge and Network Traffic Challenge: Exploring Temporal Data Sets Using a Time Series of Glyphs

Paolo Simonetto* Daniel Archambault† Faraz Zaidi* Pierre-Yves Koenig* 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 system and findings for the Badge and Network Traffic Challenge of the 2009 VAST contest. The summary starts by presenting an overview of our time series encoding of badge information and network traffic. We then present our findings and suggest that employee 30 may be of interest.

Index Terms: H.5.0 [Information Systems]: Information Interfaces and Presentation—General;

1 Introduction

In order to determine network transmissions that are the most likely candidates for leaks, we developed a visualization which encodes as much of the badge and network information as possible in a single view. This visualization was implemented using the Tulip [**?**] graph drawing libraries and software. As our visualization is fairly non-standard, we first present an overview of the encoding in Section 2. We then present our findings in section 3 and explain how we arrive at a final solution. Finally, in section 4, we present our suspect, employee 30, and why we believe this empolyee is the most likely candidate for the information leak.

2 Visualization System

Our visualization technique, shown in Figure 1, is based on a timeline view. The diagram shows, for each day, the actions of each employee. 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, 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 into the main building. Bars between two vertical lines encode the time interval between when an employee badges into the classified area to the moment the employee badges out. This period begins with a badge-in-classified event and ends with a badge-out-classified event.

The central blue bars show when the employee's computer is active. Downwardly directed circles represent transmissions. The size of the circles is proportional of the forth root of the transmission size.

A green background shows the average daily activity of an employee over the 31 days. A more saturated green indicates a higher probability that the employee is at work. Using some simple rules, we change the colour of this green background to red in order to highlight suspicious activities. The most suspicious activities occur when an employees computer is active but he or she is most likely not at their desk. These activities include:

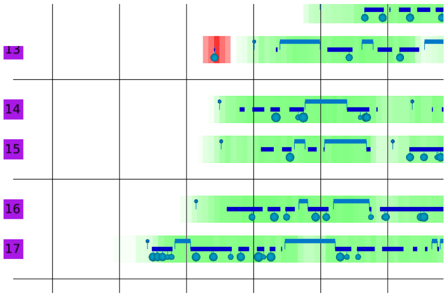

Figure 1: The first suspicious transmission found. Employee 13's computer is used for exactly one large upload on day twenty-two well before the employee usually arrives at work.

- an employee badges into the classified area but does not badge out or vice versa.
- an employee's computer is used while he or she is in the classified area.
- an employee's computer is used when the employee is not likely to be in the building.

3 Suspicious Activity

Suspicious activity was preliminarily defined by the rules described above. From these rules, we discovered several candidate transmissions that may have been involved in the leak.

3.1 Most Suspicious Activity

In our first case, Figure 1 shows a large transmission from employee 13's computer on day twenty-two well before the employee usually arrives at work. By examining the raw data, we determine that this computer activity is forty minutes before the earliest time this employee was at work in the thirty-one day period. Additionally, the badge-in-building event that was recorded for this day was typical for this employee. Notice that 13's office mate is not present at this time, giving the opportunity for a leak to be sent from this computer from a third person not assigned to this office.

In our second case, shown in Figure 2, a large transmission is sent less than a minute before employee 20 badges into the building on day twenty-nine. As the transmission precedes the badge-in-building event, it is highly unlikely that 20 was present at their desk at this time. Notice as well that 20's office mate is not present in the office, giving the opportunity for a leak to be sent from this computer from a third person not assigned to this office.

*{paolo.simonetto, faraz.zaidi, Pierre-Yves.Koenig, frederic.gilbert, phanquan, mathiaut, antoine.lambert, dubois, sicre, mathieu.brulin, vieux, }@labri.fr
†{daniel.archambault, guy.melancon}@inria.fr

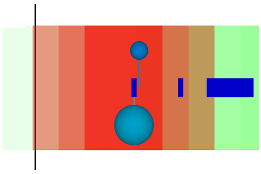

Figure 2: The second suspicious transmission. Employee 20's computer sends a large transmission on day twenty-nine just before the employee badges into the building. In this scenario, the employee is not likely at his or her desk.

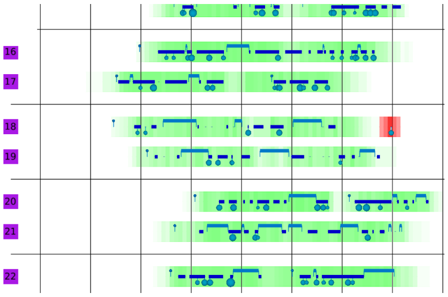

Figure 3: The third suspicious transmission. Employee 18's computer sends a large transmission when the employee is probably gone for the day.

Our third case, Figure 3, shows a large transmission sent from employee 18's computer over two hours since it was last active on day seventeen. Neither employee 18 or 19 is likely in the office at this time as the computer activity and green background indicate. Thus, it is possible for a leak to be sent from this computer by someone who is not assigned to this office.

3.2 Transmissions while in Classified Zone

Additionally, we looked for transmissions made from an employee's computer when the employee had badged into, but not out of, the classified zone. During this time, the employee's computer should not be used, because we are certain that they are away from their desk. Figure 4 shows an example, but, in reality, we found eight such cases:

- 37.170.100.31,2008-01 10T14:27:12.238,100.59.151.133,8080
- 37.170.100.16,2008-01-10T16:01:53.956,100.59.151.133,8080
- 37.170.100.16,2008-01-15T16:14:34.563,100.59.151.133,8080
- 37.170.100.41,2008-01-17T12:12:10.990,100.59.151.133,8080
- 37.170.100.56,2008-01-29T15:41:32.763,100.59.151.133,8080
- 37.170.100.41,2008-01-29T16:08:10.892,100.59.151.133,8080
- 37.170.100.52,2008-01-31T09:41:03.815,100.59.151.133,8080
- 37.170.100.15,2008-01-31T13:10:23.841,100.59.151.133,8080

3.3 Other Transmissions

Interestingly enough, all eleven of these suspicious transmissions of data are sent to the same IP address, `100.59.151.133`, and on the same port, 8080. We figured that this machine may be the machine to which the embassy leaks were uploaded. Subsequently, we highlighted all transmissions to this IP address made from the embassy and found an additional set of seven transmissions:

- 37.170.100.31,2008-01-08T17:01:33.001,100.59.151.133,8080
- 37.170.100.31,2008-01-15T17:03:29.342,100.59.151.133,8080
- 37.170.100.16,2008-01-22T17:41:55.862,100.59.151.133,8080
- 37.170.100.10,2008-01-24T09:46:34.452,100.59.151.133,8080
- 37.170.100.32,2008-01-24T10:26:31.321,100.59.151.133,8080
- 37.170.100.20,2008-01-24T17:07:34.775,100.59.151.133,8080
- 37.170.100.8,2008-01-31T16:02:44.572,100.59.151.133,8080

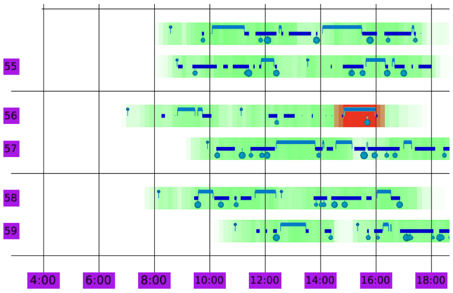

Figure 4: A case where an employee's computer is used while the employee is in the classified zone. The computer should not be used at this time, because the employee is not at his or her desk.

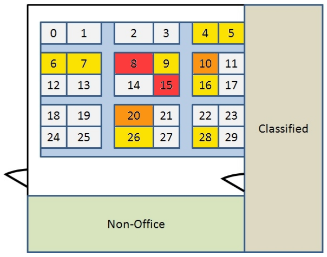

Figure 5: Embassy offices and number of suspicious transmissions made from each of them. White is zero, yellow is one, orange is two and red is three.

All of these transmissions were large and made on port 8080. In most cases, the office was probably empty with one interesting exception: employee 30 was most likely in 30/31's office when three of the suspicious transmissions were made from employee 31's computer. Thus, employee 30 seems to be a person of interest. Figure 5 plots suspicious transmissions to offices. Notice how they are clustered around office 15, which is employee 30's office.

3.4 Missing Entries in Classified Log

Finally, we found five cases where an employee either badged into the secure zone without badging out or vice versa. These events may correspond to badge error, but, if not, indicate an infraction of embasy policy. It is also interesting to note that employee 30 was involved in three of these five infractions. This observation could implicate employee 30 further as breaking this particular policy may be an attempt to collect sensitive information without being identified.

4 Conclusion

In summary, we believe that employee 30 was most likely the embassy employee who caused the leak. We have reason to believe this is the case, because most of the suspicious transmissions were made from locations near this employee's office and he or she is one of the few employees who could have sent all of these suspicious transimissions.

Acknowledgements

We thank INRIA Bordeaux Sud-Ouest Gravité project for helping support this work.