An Efficient Approach to Closing the Breach Detection Gap (BDG)

Sydney Raymond, Khalil Davis, and David Tan
Faculty Mentor: Dr. David Anyiwo
Graduate Mentor: Jerry Godwin Diabor
Bowie State University

Abstract

The breach detection gap (BDG) is a specific problem facing companies engaged in data management or retention. BDG reduction is defined as the process of limiting the time between a breach's occurrence and its detection. There are three primary factors we have targeted as potential application areas for BDG reduction methods: technology, corporate and government policy, and human error.

Introduction

The average BDG exceeds three months[1], damaging user trust in online entities and leaving users' data vulnerable to malicious actors[2]. Many factors in the current cyber environment have contributed to widening the BDG. In the interest of limiting the damage caused by intrusions, properly identifying these factors - as well as finding technological solutions that efficiently target the most significant of these factors - is essential.

Conclusions

In our literature review, we identified three potential BDG factors. Using content analysis of several case studies, we verified our hypotheses involving the association between each factor (excluding external factors) and the BDG. We then developed a program that addressed the most significant of these factors and therefore aims to reduce the BDG. The program developed by our group reduces breach detection time and is highly accurate, eliminating errors resulting from false positives.

Future Work

Future research should focus on collecting data to determine the quantitative correlation between factors and breach detection time. Organizations could contribute their individual survey data to a joint effort to understand this relationship. Additionally, further studies must be taken to more clearly analyze the relationship between external factors and breach detection time. Work should also be undertaken to measure the success of organizations that adopt our proposed strategies.

References

1. Botha, J., Eloff, M., & Swart, I. (2016). Pro-Active Data Breach Detection: Examining Accuracy and Applicability on Personal Information Detected. ICCWS 11th International Conference on Cyber Warfare and Security, 47-55.
2. Deidrich, D. (2019). Data Breaches. Georgetown Law Technology Review, 315-324.

Acknowledgements

We would like to thank Dr. Anyiwo and our graduate mentor Jerry Diabor. We would also like to thank Bowie State University, the National Science Foundation, and all of our fellow summer undergraduate research participants.

Methodology

Table 1 describes our hypotheses.

Hypothesis #1	H1a: An association exists between technology and the BDG.
	H1b: There is a negative correlation between correct technology application and the BDG.
Hypothesis #2	H2a: An association exists between corporate and government policies and the BDG.
	H2b: There is a negative correlation between effective policies and the BDG.
Hypothesis #3	H3a: An association exists between human error and the BDG.
	H3b: There is a positive correlation between human error and the BDG.
Hypothesis #4	H4a: The external environment influences the BDG.
	H4b: The internal influence always acts to delay breach detection.

Table 2 describes our hypothesis verification.

Hypothesis #1	H1a: Verified via content analysis
	H1b: Requires further analysis
Hypothesis #2	H2a: Verified via content analysis
	H2b: Requires further analysis
Hypothesis #3	H3a: Verified via content analysis
	H3b: Requires further analysis
Hypothesis #4	H4a: Requires further analysis
	H4b: Requires further analysis

Fig. 1 illustrates the data breach process.

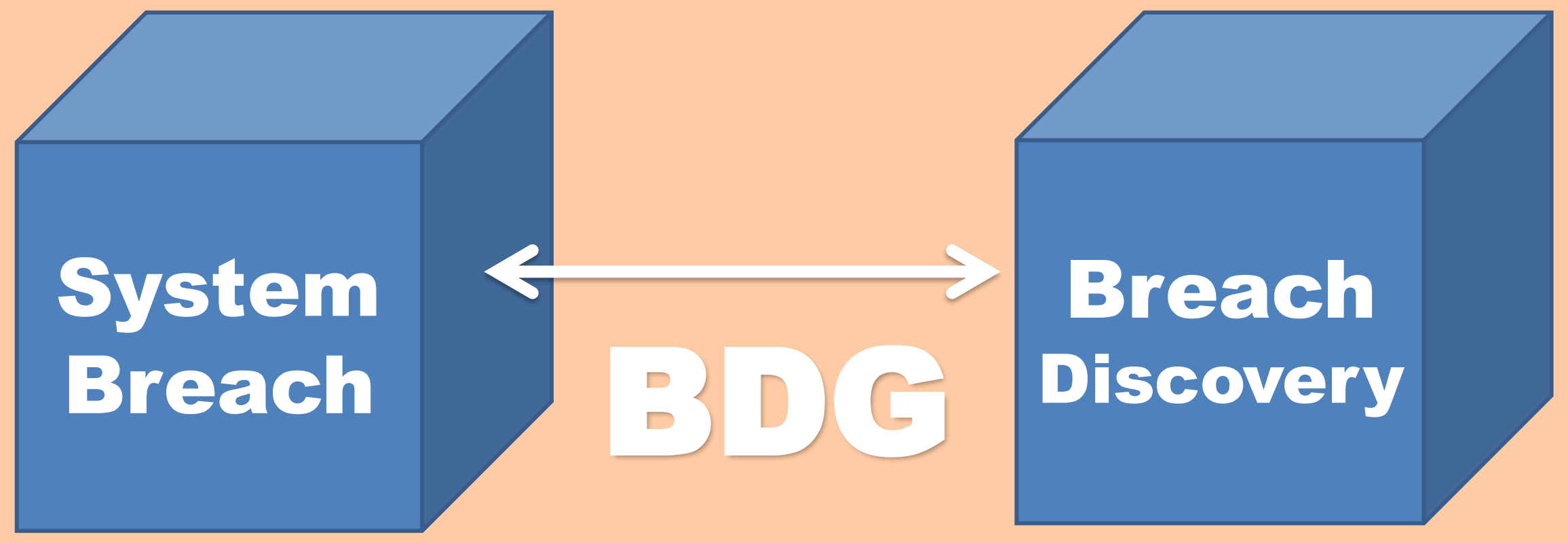
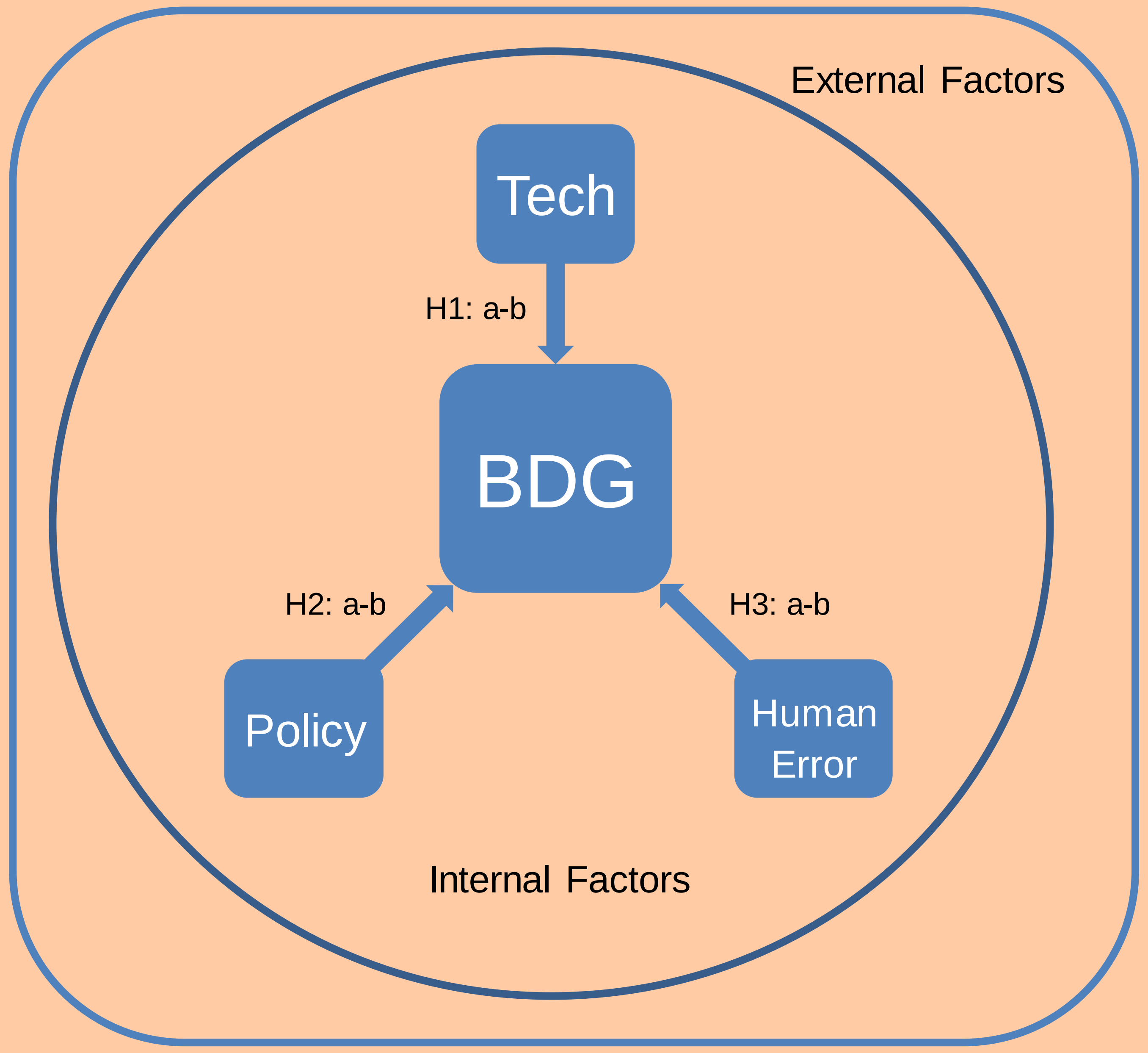
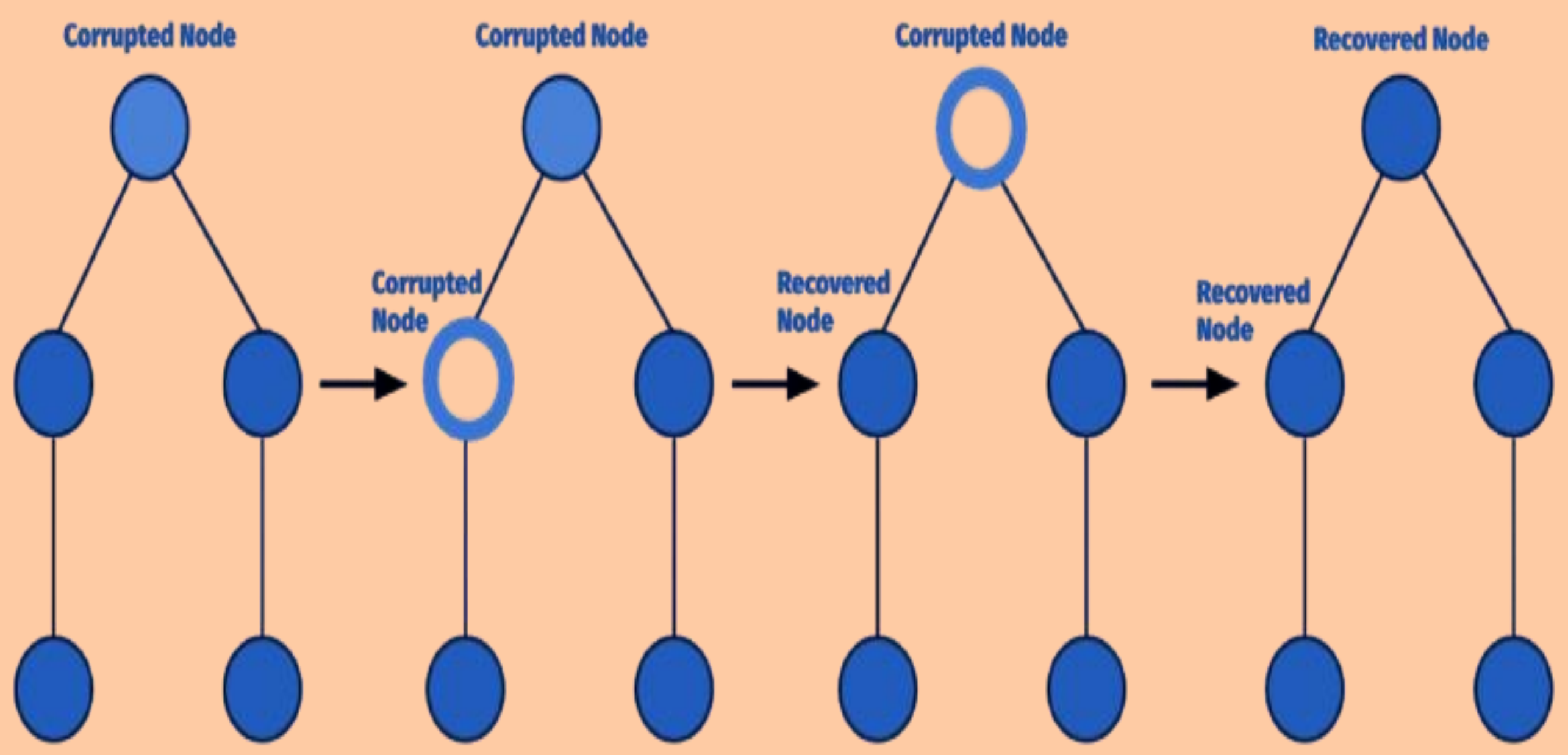


Fig. 2 shows the relationship between BDG factors.

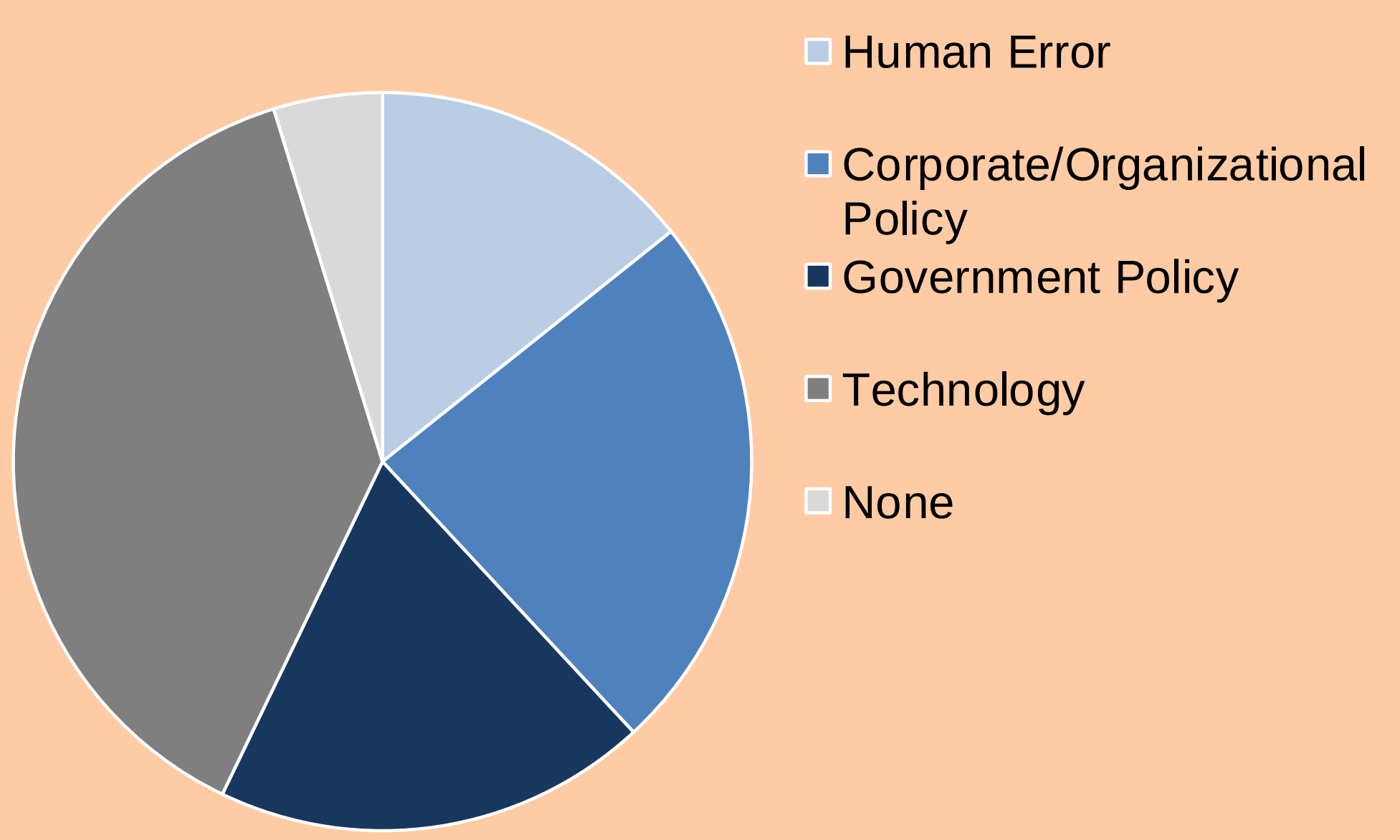


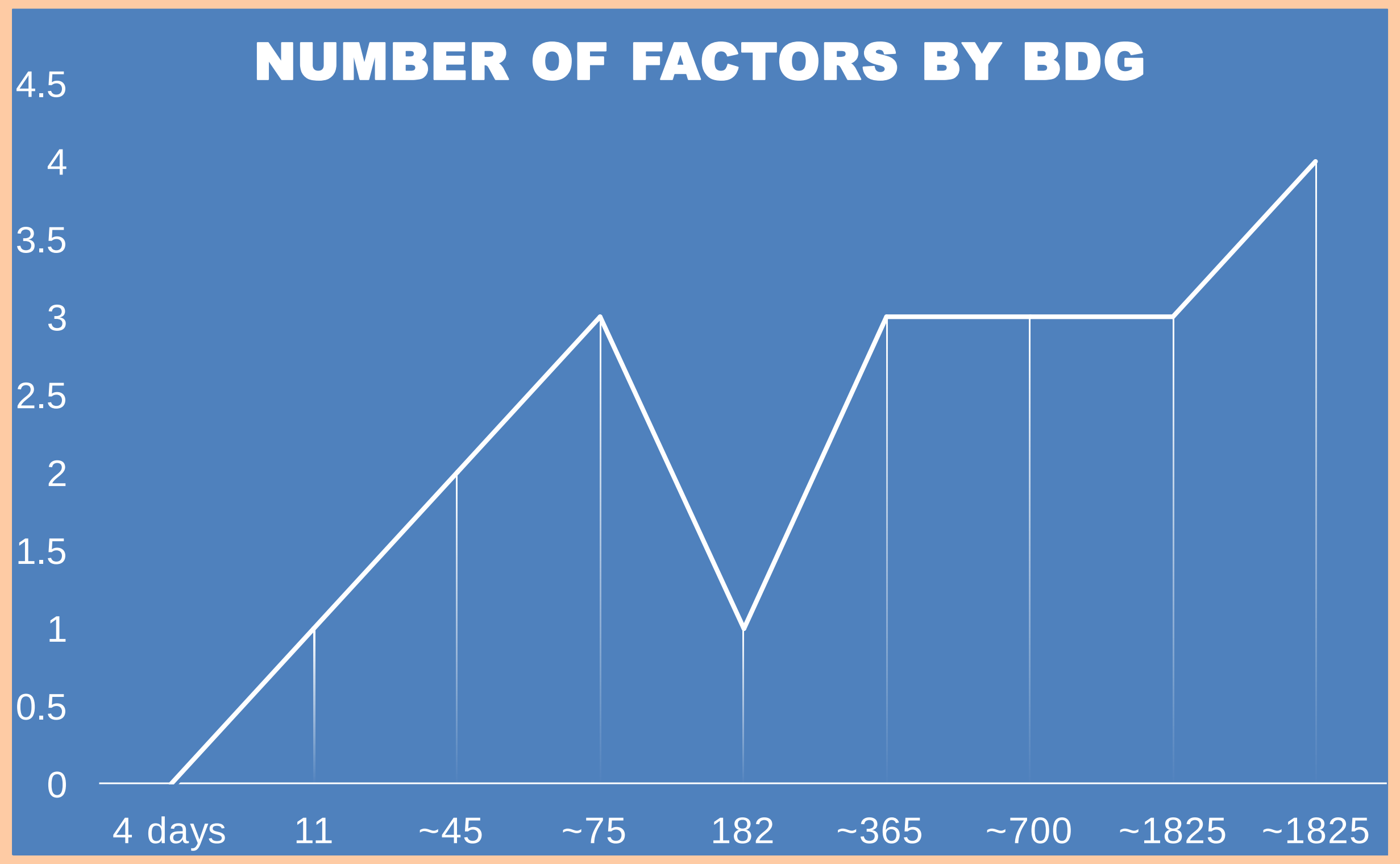
Results

Fig. 3 demonstrates the program concept.

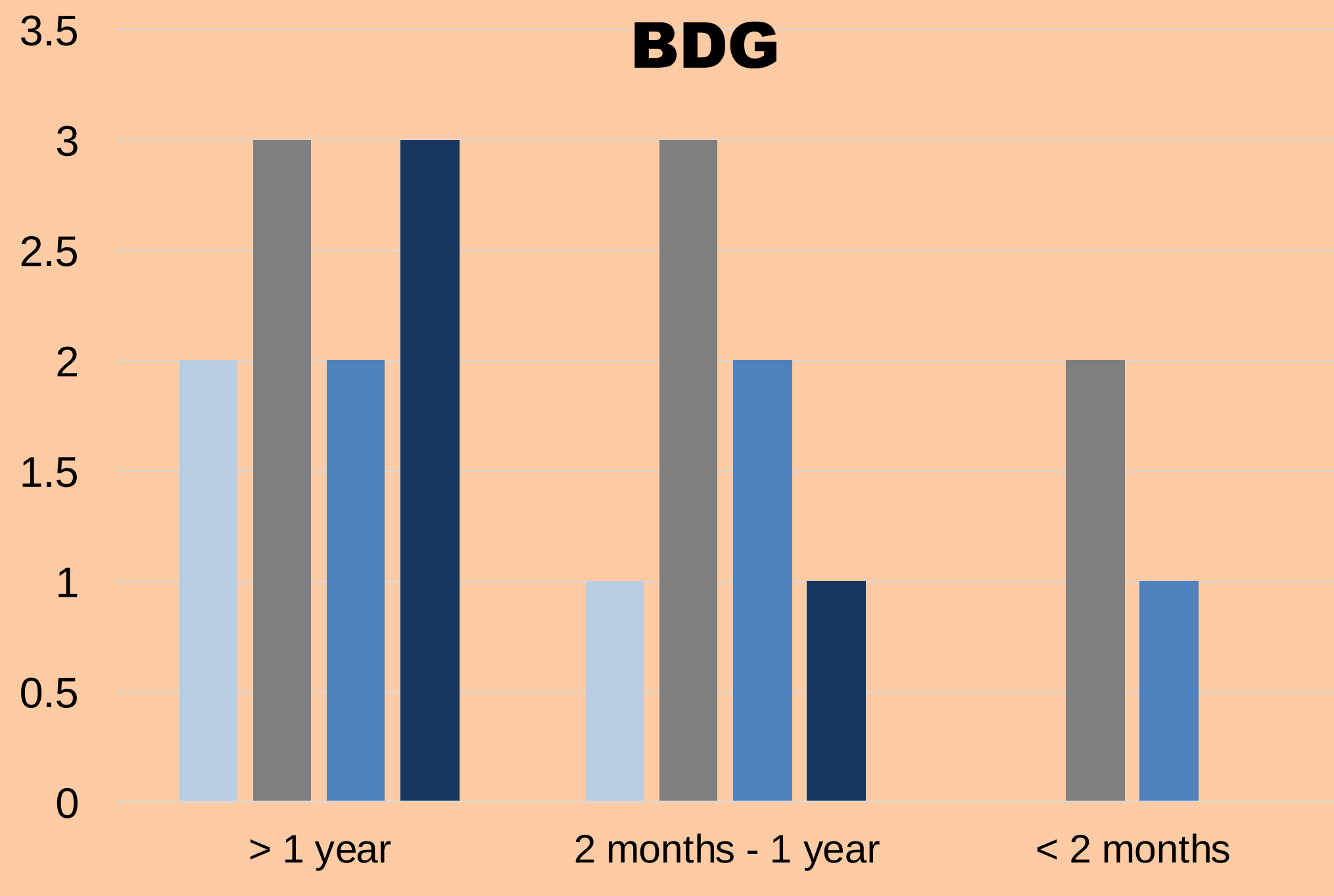


BDG Factors

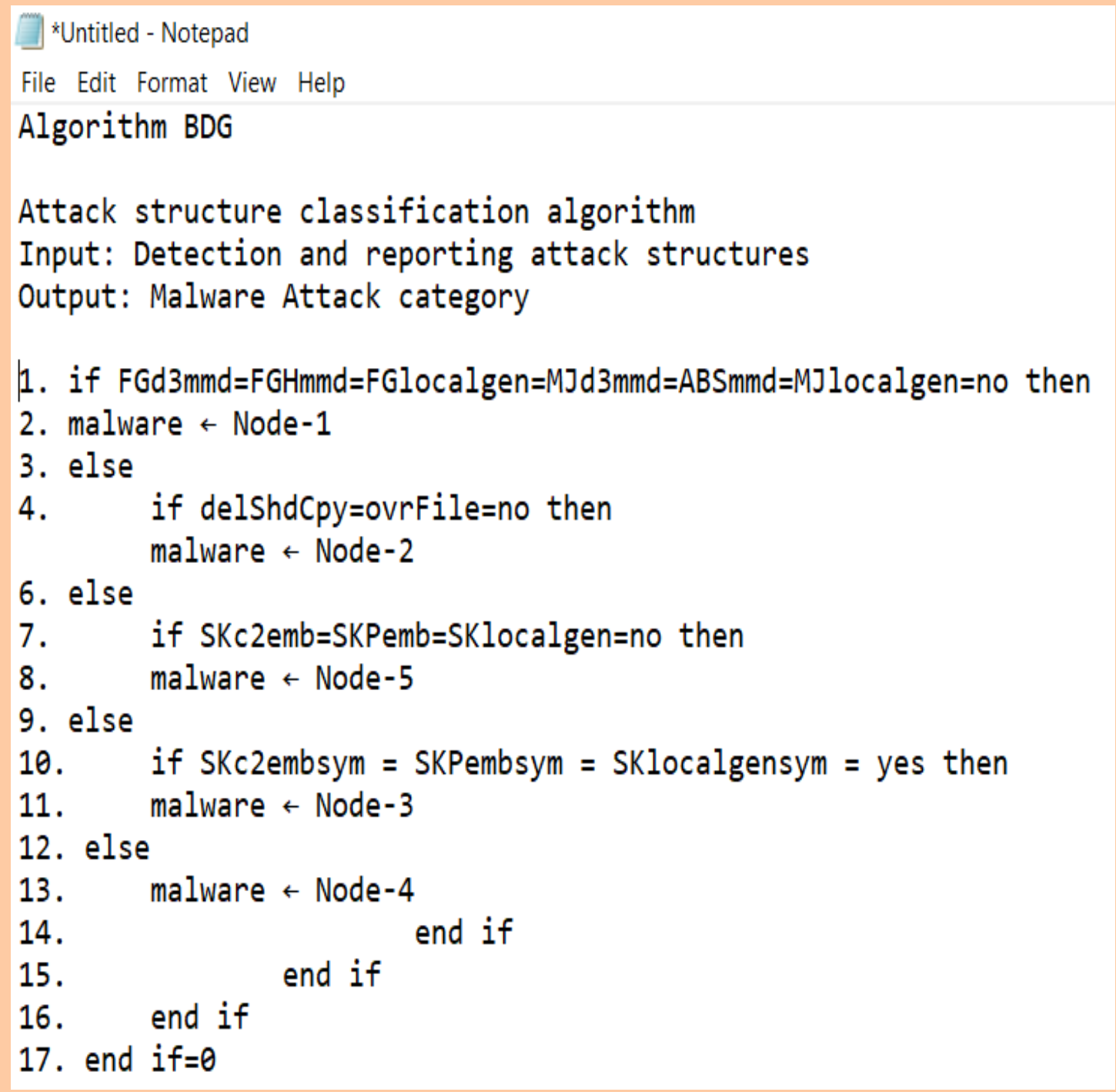




Factor Frequency Grouped By BDG



Algorithm - BDG

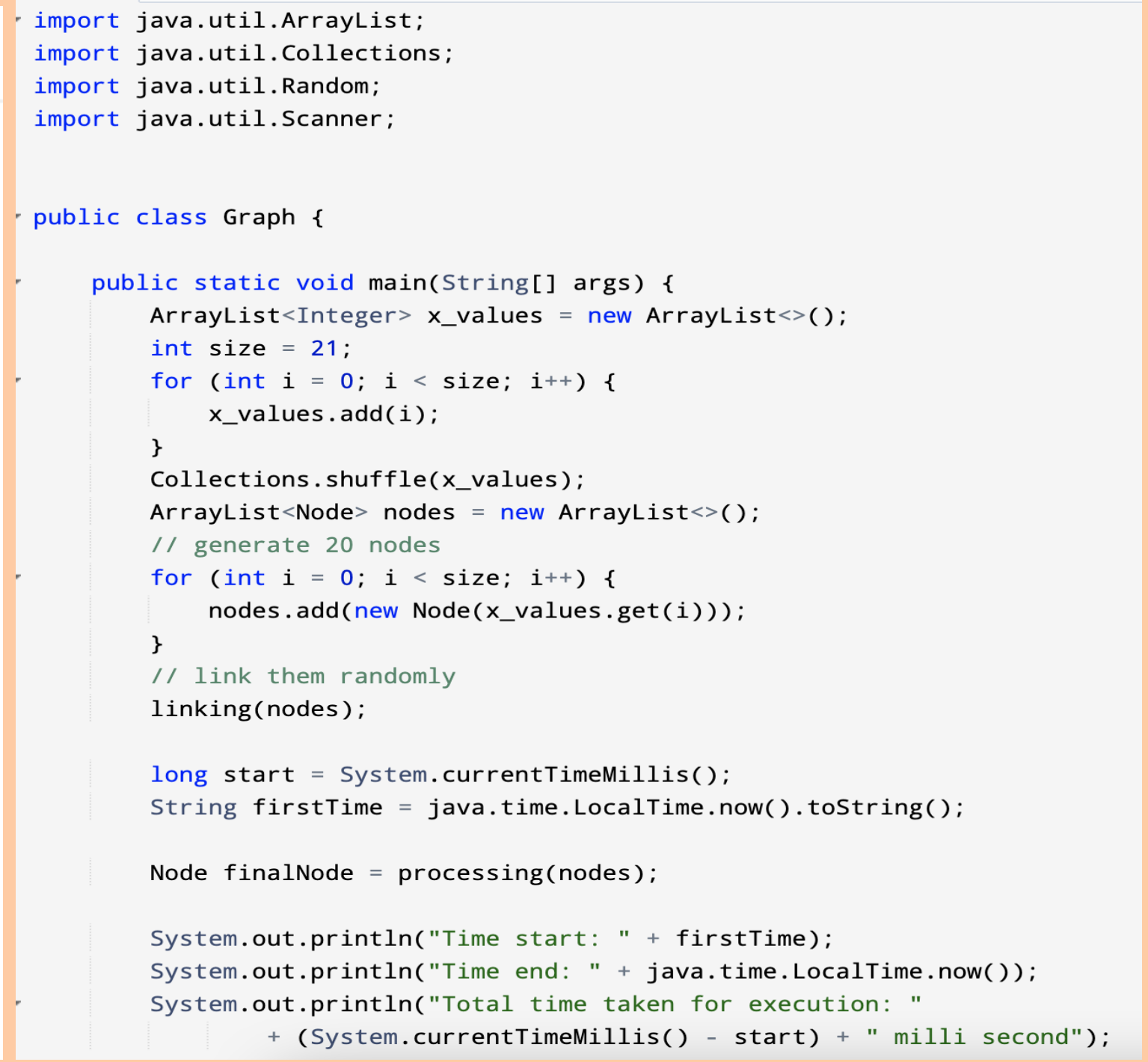


```
Algorithm BDG

Attack structure classification algorithm
Input: Detection and reporting attack structures
Output: Malware Attack category

1. if FGd3mmd=FGHmmd=FGlocalgen=MJd3mmd=ABSmmd=MJlocalgen=no then
2. malware ← Node-1
3. else
4.      if delShdCpy=ovrFile=no then
        malware ← Node-2
6. else
7.      if SKc2emb=SKPemb=SKlocalgen=no then
8.      malware ← Node-5
9. else
10.     if SKc2embsym = SKPembsym = SKlocalgensym = yes then
11.     malware ← Node-3
12. else
13.     malware ← Node-4
14.                     end if
15.             end if
16.     end if
17. end if=0
```

Sample Code - BDG



```java
import java.util.ArrayList;
import java.util.Collections;
import java.util.Random;
import java.util.Scanner;


public class Graph {

    public static void main(String[] args) {
        ArrayList<Integer> x_values = new ArrayList<>();
        int size = 21;
        for (int i = 0; i < size; i++) {
            x_values.add(i);
        }
        Collections.shuffle(x_values);
        ArrayList<Node> nodes = new ArrayList<>();
        // generate 20 nodes
        for (int i = 0; i < size; i++) {
            nodes.add(new Node(x_values.get(i)));
        }
        // link them randomly
        linking(nodes);

        long start = System.currentTimeMillis();
        String firstTime = java.time.LocalTime.now().toString();

        Node finalNode = processing(nodes);

        System.out.println("Time start: " + firstTime);
        System.out.println("Time end: " + java.time.LocalTime.now());
        System.out.println("Total time taken for execution: "
                + (System.currentTimeMillis() - start) + " milli second");
```

Table 4 depicts average stabilization time as the node number increases to 200.

Node Number	20	40	60	80	100	120	140	160	180	200
	26579	55572	72356	110259	132021	153226	210065	221569	243962	277231
	4207	41235	55465	63569	124436	110363	196362	206654	223651	256213
	3477	22578	52365	72569	105569	123554	153265	215565	241623	263316
	8791	32152	12258	82246	113659	151235	206543	213326	235663	271621
	9274	38426	36542	100698	98625	136854	192236	196354	243316	243165
	80	42315	39569	95562	105378	110369	206321	205698	233656	269563
	853	21589	71125	73256	121369	146559	194563	220364	237128	275363
	819	37856	46589	83465	115639	142396	186336	199856	231236	265436
	17971	51582	44569	103356	126963	152646	201136	200656	229633	256312
	29	26539	25349	756233	110639	134886	204563	211566	240361	273165
Total	72080	369844	456187	1541213	1154298	1362088	1951390	2091608	2360229	2651385
Average	7208	36984.4	45618.7	154121.3	115429.8	136208.8	195139	209160.8	236022.9	265138.5